

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 10, 2004, 17:10:22 ; Search time 10.9751 Seconds
(without alignments)
7635.278 Million cell updates/sec

Title: US-10-001-857-41

Perfect score: 151

Sequence: 1 ccgcggcgagtgactaa.....gcaaatctctcaagaatg 151

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listed first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*

2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*

3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*

4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*

5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	31	20.5	1689	4	US-09-489-039A-840
C 2	30.2	20.0	4421	4	US-08-952-674-1
C 3	29.4	19.5	2348	3	US-08-480-173A-42
C 4	29.4	19.5	2348	3	US-08-484-408A-42
C 5	29.4	19.5	3182	4	US-08-890-735C-1
C 6	29.4	19.5	3182	4	US-10-104-966-14
C 7	29.4	19.5	5618	3	US-08-799-569-1
C 8	29.4	19.5	5618	4	US-09-570-546-1
C 9	29.4	19.5	5618	4	US-09-146-072A-1
C 10	29.4	19.2	176373	3	US-09-128-155-17
C 11	28.6	18.9	3215	4	US-09-719-528A-1
C 12	28	18.5	786431	4	US-09-751-389-3
C 13	27.8	18.4	49312	4	US-09-671-317-485
C 14	27.4	18.1	1664976	4	US-08-916-421B-1
C 15	27.2	18.0	595	4	US-08-976-259-87
C 16	26.8	17.7	1740	4	US-09-328-352-999
C 17	26.8	17.7	168575	4	US-09-426-290-1
C 18	26.6	17.6	501	4	US-09-621-976-2040
C 19	26.6	17.6	2089	4	US-09-620-312D-303
C 20	26.6	17.6	3069	4	US-09-514-907A-5
C 21	26.6	17.6	3069	4	US-09-896-994-5
C 22	26.6	17.6	392000	4	US-10-027-983-11
C 23	26.6	17.6	580073	4	US-08-545-528D-1
C 24	26.4	17.5	577	3	US-09-328-111-606
C 25	26.4	17.5	1671	4	US-09-545-686-1
C 26	26.2	17.4	2111	4	US-10-096-571-11
C 27	26.2	17.4	2111	4	US-10-096-571-13

C 28	26.2	17.4	2841	2	US-08-452-075-1	Sequence 1, Appli
C 29	26.2	17.4	2841	3	US-09-231-061-1	Sequence 1, Appli
C 30	26.2	17.4	2841	4	US-09-011-762-5	Sequence 5, Appli
C 31	26	17.2	312	4	US-09-313-294A-7377	Sequence 7377, Ap
C 32	26	17.2	779	4	US-08-956-171E-1019	Sequence 1019, Ap
C 33	26	17.2	1449	4	US-08-956-171E-592	Sequence 592, App
C 34	26	17.2	41708	4	US-09-470-512A-3	Sequence 3, Appli
C 35	26	17.2	392000	4	US-10-027-983-11	Sequence 11, Appli
C 36	26	17.2	1830121	4	US-09-557-884-1	Sequence 1, Appli
C 37	26	17.2	1830121	4	US-09-643-990A-1	Sequence 1035, Ap
C 38	25.8	17.1	618	4	US-09-489-039A-1035	Sequence 8, Appli
C 39	25.8	17.1	5108	4	US-09-306-446C-8	Sequence 5, Appli
C 40	25.8	17.1	5889	1	US-08-463-092B-5	Sequence 5, Appli
C 41	25.8	17.1	5889	2	US-08-462-109A-5	Sequence 5, Appli
C 42	25.8	17.1	5889	2	US-08-460-907B-5	Sequence 5, Appli
C 43	25.8	17.1	5889	3	US-08-463-179A-5	Sequence 5, Appli
C 44	25.8	17.1	5889	3	US-08-461-384B-5	Sequence 5, Appli
C 45	25.8	17.1	580073	4	US-08-545-528D-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-489-039A-840/c
; Sequence 840, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; CURRENT FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 840

; LENGTH: 1689

; TYPE: DNA

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-840

Query Match 20.5%; Score 31; DB 4; Length 1689;
Best Local Similarity 68.3%; Pred. No. 0.24;
Matches 43; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY	28	AAATGTCCTTTGGGGTTCCTGTTTCAACAGCATGTTGTAAGCGCCGATCAACCTTC	87
Db	1607	ACATGAAGCAGGTATGCCCTCGCTGGAAGCATGGTGTGAGTCCGCTTCACTGG	1548
QY	88	TCT 90	
Db	1547	TCT 1545	

RESULT 2

US-08-952-674-1/c

; Sequence 1, Application US/08952674

; Patent No. 6623951

; GENERAL INFORMATION:

; APPLICANT: Hofschneider, Peter

; APPLICANT: Habenberger, Peter

; APPLICANT: Ludwig, Weiss

; TITLE OF INVENTION: HBV Vectors and Cells for Producing Same

; FILE REFERENCE: HBV Vectors

; CURRENT APPLICATION NUMBER: US/08/952,674

; CURRENT FILING DATE: 1998-03-02

; EARLIER APPLICATION NUMBER: PCT/DE96/00807

; EARLIER FILING DATE: 1996-05-09

; NUMBER OF SEQ ID NOS: 1

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

Query	Match	Score	DB 1	DB 2	DB 3	DB 4	DB 5	DB 6	DB 7	DB 8	DB 9	DB 10	DB 11	DB 12	DB 13	DB 14	DB 15	DB 16	DB 17	DB 18	DB 19	DB 20	DB 21	DB 22	DB 23	DB 24	DB 25	DB 26	DB 27	DB 28	DB 29	DB 30	DB 31	DB 32	DB 33	DB 34	DB 35	DB 36	DB 37	DB 38	DB 39	DB 40	DB 41	DB 42	DB 43	DB 44	DB 45	DB 46	DB 47	DB 48	DB 49	DB 50	DB 51	DB 52	DB 53	DB 54	DB 55	DB 56	DB 57	DB 58	DB 59	DB 60	DB 61	DB 62	DB 63	DB 64	DB 65	DB 66	DB 67	DB 68	DB 69	DB 70	DB 71	DB 72	DB 73	DB 74	DB 75	DB 76	DB 77	DB 78	DB 79	DB 80	DB 81	DB 82	DB 83	DB 84	DB 85	DB 86	DB 87	DB 88	DB 89	DB 90	DB 91	DB 92	DB 93	DB 94	DB 95	DB 96	DB 97	DB 98	DB 99	DB 100	DB 101	DB 102	DB 103	DB 104	DB 105	DB 106	DB 107	DB 108	DB 109	DB 110	DB 111	DB 112	DB 113	DB 114	DB 115	DB 116	DB 117	DB 118	DB 119	DB 120	DB 121	DB 122	DB 123	DB 124	DB 125	DB 126	DB 127	DB 128	DB 129	DB 130	DB 131	DB 132	DB 133	DB 134	DB 135	DB 136	DB 137	DB 138	DB 139	DB 140	DB 141	DB 142	DB 143	DB 144	DB 145	DB 146	DB 147	DB 148	DB 149	DB 150	DB 151	DB 152	DB 153	DB 154	DB 155	DB 156	DB 157	DB 158	DB 159	DB 160	DB 161	DB 162	DB 163	DB 164	DB 165	DB 166	DB 167	DB 168	DB 169	DB 170	DB 171	DB 172	DB 173	DB 174	DB 175	DB 176	DB 177	DB 178	DB 179	DB 180	DB 181	DB 182	DB 183	DB 184	DB 185	DB 186	DB 187	DB 188	DB 189	DB 190	DB 191	DB 192	DB 193	DB 194	DB 195	DB 196	DB 197	DB 198	DB 199	DB 200	DB 201	DB 202	DB 203	DB 204	DB 205	DB 206	DB 207	DB 208	DB 209	DB 210	DB 211	DB 212	DB 213	DB 214	DB 215	DB 216	DB 217	DB 218	DB 219	DB 220	DB 221	DB 222	DB 223	DB 224	DB 225	DB 226	DB 227	DB 228	DB 229	DB 230	DB 231	DB 232	DB 233	DB 234	DB 235	DB 236	DB 237	DB 238	DB 239	DB 240	DB 241	DB 242	DB 243	DB 244	DB 245	DB 246	DB 247	DB 248	DB 249	DB 250	DB 251	DB 252	DB 253	DB 254	DB 255	DB 256	DB 257	DB 258	DB 259	DB 260	DB 261	DB 262	DB 263	DB 264	DB 265	DB 266	DB 267	DB 268	DB 269	DB 270	DB 271	DB 272	DB 273	DB 274	DB 275	DB 276	DB 277	DB 278	DB 279	DB 280	DB 281	DB 282	DB 283	DB 284	DB 285	DB 286	DB 287	DB 288	DB 289	DB 290	DB 291	DB 292	DB 293	DB 294	DB 295	DB 296	DB 297	DB 298	DB 299	DB 300	DB 301	DB 302	DB 303	DB 304	DB 305	DB 306	DB 307	DB 308	DB 309	DB 310	DB 311	DB 312	DB 313	DB 314	DB 315	DB 316	DB 317	DB 318	DB 319	DB 320	DB 321	DB 322	DB 323	DB 324	DB 325	DB 326	DB 327	DB 328	DB 329	DB 330	DB 331	DB 332	DB 333	DB 334	DB 335	DB 336	DB 337	DB 338	DB 339	DB 340	DB 341	DB 342	DB 343	DB 344	DB 345	DB 346	DB 347	DB 348	DB 349	DB 350	DB 351	DB 352	DB 353	DB 354	DB 355	DB 356	DB 357	DB 358	DB 359	DB 360	DB 361	DB 362	DB 363	DB 364	DB 365	DB 366	DB 367	DB 368	DB 369	DB 370	DB 371	DB 372	DB 373	DB 374	DB 375	DB 376	DB 377	DB 378	DB 379	
-------	-------	-------	------	------	------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--

Db 2094 AACTCCTCCAGTCTTTAAACAAACAGTCTTTGAAGTATGCTCAAGGTG 2044

RESULT 5
US-08-890-735C-1/c
; Sequence 1, Application US/08890735C
; Patent No. 6518014
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: HEPADNAVIRUS CORES
; FILE REFERENCE: DC44A
; CURRENT APPLICATION NUMBER: US/08/890,735C
; CURRENT FILING DATE: 1997-07-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 3182
; TYPE: DNA
; ORGANISM: Hepatitis B Virus
US-08-890-735C-1

Query Match 19.5%; Score 29.4; DB 4; Length 3182;
Best Local Similarity 54.1%; Pred. No. 1.2;
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 22 CAGGCCAAATGTCCTTTGGGGTTCCTGTTTCAACAGCATGTTGTGAAGCGCGCATCA 81
Db 1803 CAGACCAATTTATGCTTACAGCTCTCTAGTACAAAGACCTTTAACCTAATCTCTCCGCC 1744

QY 82 ACCTTCTGCTCTATTAAATAAATGTCATAAACTATCCTGCAAGGTGG 132
Db 1743 AACTCCTCCAGTCTTTAAACAAACAGTCTTTGAAGTATGCTCAAGGTG 1693

RESULT 6
US-10-104-966-14/c
; Sequence 14, Application US/10104966
; Patent No. 6680059
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: TRIPEP.23AUSC1
; CURRENT APPLICATION NUMBER: US/10/104,966
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 3182
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis B virus sequence
US-10-104-966-14

Query Match 19.5%; Score 29.4; DB 4; Length 3182;
Best Local Similarity 54.1%; Pred. No. 1.2;
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 22 CAGGCCAAATGTCCTTTGGGGTTCCTGTTTCAACAGCATGTTGTGAAGCGCGCATCA 81
Db 1803 CAGACCAATTTATGCTTACAGCTCTCTAGTACAAAGACCTTTAACCTAATCTCTCCGCC 1744

QY 82 ACCTTCTGCTCTATTAAATAAATGTCATAAACTATCCTGCAAGGTGG 132
Db 1743 AACTCCTCCAGTCTTTAAACAAACAGTCTTTGAAGTATGCTCAAGGTG 1693

RESULT 7
US-08-799-569-1/c
; Sequence 1, Application US/08799569
; Patent No. 6133244
; GENERAL INFORMATION:
; APPLICANT: Michel, Marie-Louise
; APPLICANT: Mancine, Maryline
; TITLE OF INVENTION: Nucleotide Vector, Composition
; TITLE OF INVENTION: Containing Such Vector, and Vaccine for Immunization
; TITLE OF INVENTION: Against Hepatitis
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/799,569
; FILING DATE: 12-FEB-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/706,337
; FILING DATE: 30-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/633,821
; FILING DATE: 22-APR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 94/00483
; FILING DATE: 27-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0128-01000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5618 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-799-569-1

Query Match 19.5%; Score 29.4; DB 3; Length 5618;
Best Local Similarity 54.1%; Pred. No. 1.5;
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 22 CAGGCCAAATGTCCTTTGGGGTTCCTGTTTCAACAGCATGTTGTGAAGCGCGCATCA 81
Db 2585 CAGACCAATTTATGCTTACAGCTCTCTAGTACAAAGACCTTTAACCTAATCTCTCCGCC 2526

QY 82 ACCTTCTGCTCTATTAAATAAATGTCATAAACTATCCTGCAAGGTGG 132
Db 2525 AACTCCTCCAGTCTTTAAACAAACAGTCTTTGAAGTATGCTCAAGGTG 2475

RESULT 8
US-09-570-546-1/c
; Sequence 1, Application US/09570546
; Patent No. 6429201
; GENERAL INFORMATION:
; APPLICANT: Michel, Marie-Louise

```

; CURRENT FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: US 08/633,821
; PRIOR FILING DATE: 1996-08-02
; PRIOR APPLICATION NUMBER: PCT/FR94/000483
; PRIOR FILING DATE: 1994-04-27
; PRIOR APPLICATION NUMBER: FR 93/12659
; PRIOR FILING DATE: 1993-10-22
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 5618
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Synthetic Sequence
US-09-146-072A-1

Query Match      19.5%; Score 29.4; DB 4; Length 5618;
Best Local Similarity 54.1%; Pred. No.1.5;
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY      22 CAGCCCAAATGTTGCCCTTTGGGHTTCTGTTTCAACAGCATGGTGTGAAGCGCGCATCA 81
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2585 CAGACCAATTATGCTCTACAGCCTCTAGTACAAAGACCTTTAACTTAATCTCTCTCCCC 2526

QY      82 ACCTTCTCTGCTATTAAATAAATGTCATAACTCATCTCTCAAGGTGG 132
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2525 AACTCTCCAGTCTTTAAACAAACAGTCTTTGAAGTATGCTCAAGGTGG 2475

RESULT 10
US-09-128-155-17/c
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 176373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(176373)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17

Query Match      19.2%; Score 29; DB 3; Length 176373;
Best Local Similarity 53.6%; Pred. No.10;
Matches 59; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY      4 CCGCGCAGGTACCTAAACAGCGCCAAATGTCCTTTGGGTTCTGTTTCAACAGCATG 63
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8396 CCTGTCAATACTAAGGTGTCAAAATTGNCATAGGGGTACTTTACTATAGCGCA 8337

QY      64 GTGTGAAGCGCGCATCAACCTTCTGCTATTAAAAATAAATATGCATA 113
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8336 GCCAAACAGAATACTTTTGCTTTTTCATCTTAAATTAATGTTAGA 8287

RESULT 11
US-09-719-528A-1/c
; Sequence 1, Application US/09719528A

```



```
; Patent No. 6558675
; GENERAL INFORMATION:
; APPLICANT: Oon, Chong Jin
; Lim, Gek Keow
; Zhao, Yi
; Chen, Wei Ning
; TITLE OF INVENTION: A MUTANT HUMAN HEPATITIS B VIRAL STRAIN AND
; USES THEREOF
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ladas & Parry
; STREET: 26 West 61 Street
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10023
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/719,528A
; FILING DATE: 30-Apr-2001
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/SG98/00046
; FILING DATE: 19-JAN-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Mass, Clifford J.
; REGISTRATION NUMBER: 30,086
; REFERENCE/DOCKET NUMBER: U-013109-7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 708-1800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3215 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-719-528A-1
Query Match 18.9%; Score 28.6; DB 4; Length 3215;
Best Local Similarity 53.0%; Pred. No. 2.3;
Matches 61; Conservative 0; Mismatches 54; Indels 0; Gaps 0;
QY 18 TAAACAGCCAAATGTTGCCCTTTGGGCTTCTGTTTCAACAGCATGTTGTGAAGCGCGC 77
Db 1805 TGAACAGACCAATTTATGCCCTTACAGCTCCTAGTACATAATCTTTAACTCTCTC 1746
QY 78 ATCAACCTTCTCTGCCTATTAAATAAAATGTCATAAACTCATCTCGAAGGTGG 132
Db 1745 CCCCACTCTCCAGTCTTTAAACACACAGTCTTTGAAGTATGCTCAAGGTG 1691
RESULT 12
US-09-751-389-3
; Sequence 3, Application US/09751389
; Patent No. 6630334
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; THEREOF
; FILE REFERENCE: C1001067
; CURRENT APPLICATION NUMBER: US/09/751,389
; CURRENT FILING DATE: 2001-01-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 786431
; TYPE: DNA
```

```
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(786431)
; OTHER INFORMATION: n = A,T,C or G
US-09-751-389-3
Query Match 18.5%; Score 28; DB 4; Length 786431;
Best Local Similarity 66.7%; Pred. No. 45;
Matches 40; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
QY 44 GTTCCTGTTTCAACAGCATGTTGTGAAGCGCGCATCAACCTTCTCTGCTATATAATA 103
Db 15775 GTTACTGTTACAGGACATATTCCTTAAGTAGTTTCAATCTGTCTGCTATTAACTA 157834
RESULT 13
US-09-671-317-485
; Sequence 485, Application US/09671317
; Patent No. 6528260
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS RELATED TO GENES INVOLVED IN DRUG METABOLISM
; FILE REFERENCE: 62.US3.CIP
; CURRENT APPLICATION NUMBER: US/09/671,317
; CURRENT FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US 09/536,178
; PRIOR FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: PCT/IB00/00403
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 60/126,269
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/131,961
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 977
; SOFTWARE: Patent.pm
; SEQ ID NO 485
; LENGTH: 49312
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 5466..7466
; OTHER INFORMATION: 5'regulatory region
; NAME/KEY: exon
; LOCATION: 7467..7725
; OTHER INFORMATION: exon 1
; NAME/KEY: exon
; LOCATION: 20256..20355
; OTHER INFORMATION: exon 2
; NAME/KEY: exon
; LOCATION: 36905..36975
; OTHER INFORMATION: exon 3
; NAME/KEY: exon
; LOCATION: 45167..45248
; OTHER INFORMATION: exon 4
; NAME/KEY: exon
; LOCATION: 45728..45965
; OTHER INFORMATION: exon 5
; NAME/KEY: misc feature
; LOCATION: 45966..49312
; OTHER INFORMATION: 3'regulatory region
; NAME/KEY: allele
; LOCATION: 7564
; OTHER INFORMATION: 10-286-289 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 7619
; OTHER INFORMATION: 10-286-345 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 7649
```

```

; OTHER INFORMATION: 10-286-375 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 17258
; OTHER INFORMATION: 12-425-57 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 21590
; OTHER INFORMATION: 12-421-135 : insertion of T
; NAME/KEY: allele
; LOCATION: 21595
; OTHER INFORMATION: 12-421-140 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 36971
; OTHER INFORMATION: 10-523-232 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 45214
; OTHER INFORMATION: 10-289-201 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 45741
; OTHER INFORMATION: 10-290-37 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 46029
; OTHER INFORMATION: 10-290-326 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 46032
; OTHER INFORMATION: 10-290-328 : deletion of G
; NAME/KEY: primer_bind
; LOCATION: 7276..7294
; OTHER INFORMATION: 10-286.pu
; NAME/KEY: primer_bind
; LOCATION: 7676..7694
; OTHER INFORMATION: 10-286.rp complement
; NAME/KEY: primer_bind
; LOCATION: 16839..16856
; OTHER INFORMATION: 12-425.rp
; NAME/KEY: primer_bind
; LOCATION: 17297..17314
; OTHER INFORMATION: 12-425.pu complement
; NAME/KEY: primer_bind
; LOCATION: 21456..21474
; OTHER INFORMATION: 12-421.pu
; NAME/KEY: primer_bind
; LOCATION: 2186..2196
; OTHER INFORMATION: 12-421.rp complement
; NAME/KEY: primer_bind
; LOCATION: 36740..36758
; OTHER INFORMATION: 10-523.pu
; NAME/KEY: primer_bind
; LOCATION: 36997..37015
; OTHER INFORMATION: 10-523.rp complement
; NAME/KEY: primer_bind
; LOCATION: 45020..45037
; OTHER INFORMATION: 10-289.pu
; NAME/KEY: primer_bind
; LOCATION: 45413..45432
; OTHER INFORMATION: 10-289.rp complement
; NAME/KEY: primer_bind
; LOCATION: 45705..45724
; OTHER INFORMATION: 10-290.pu
; NAME/KEY: primer_bind
; LOCATION: 46104..46123
; OTHER INFORMATION: 10-290.rp complement
; NAME/KEY: primer_bind
; LOCATION: 7545..7563
; OTHER INFORMATION: 10-286-289.mis
; NAME/KEY: primer_bind
; LOCATION: 7565..7583
; OTHER INFORMATION: 10-286-289.mis complement
; NAME/KEY: primer_bind
; LOCATION: 7600..7618
; OTHER INFORMATION: 10-286-345.mis
; NAME/KEY: primer_bind
; LOCATION: 7620..7638
; OTHER INFORMATION: 10-286-345.mis complement

```

```

; NAME/KEY: primer_bind
; LOCATION: 7630..7648
; OTHER INFORMATION: 10-286-375.mis
; NAME/KEY: primer_bind
; LOCATION: 7650..7668
; OTHER INFORMATION: 10-286-375.mis complement
; NAME/KEY: primer_bind
; LOCATION: 17239..17257
; OTHER INFORMATION: 12-425-57.mis
; NAME/KEY: primer_bind
; LOCATION: 17259..17277
; OTHER INFORMATION: 12-425-57.mis complement
; NAME/KEY: primer_bind
; LOCATION: 21576..21594
; OTHER INFORMATION: 12-421-140.mis
; NAME/KEY: primer_bind
; LOCATION: 21596..21614
; OTHER INFORMATION: 12-421-140.mis complement
; NAME/KEY: primer_bind
; LOCATION: 36952..36970
; OTHER INFORMATION: 10-523-232.mis
; NAME/KEY: primer_bind
; LOCATION: 36972..36990
; OTHER INFORMATION: 10-523-232.mis complement
; NAME/KEY: primer_bind
; LOCATION: 45195..45213
; OTHER INFORMATION: 10-289-201.mis
; NAME/KEY: primer_bind
; LOCATION: 45215..45233
; OTHER INFORMATION: 10-289-201.mis complement
; NAME/KEY: primer_bind
; LOCATION: 45722..45740
; OTHER INFORMATION: 10-290-37.mis
; NAME/KEY: primer_bind
; LOCATION: 45742..45760
; OTHER INFORMATION: 10-290-37.mis complement
; NAME/KEY: primer_bind
; LOCATION: 46010..46028
; OTHER INFORMATION: 10-290-326.mis
; NAME/KEY: primer_bind
; LOCATION: 46030..46048
; OTHER INFORMATION: 10-290-326.mis complement
; NAME/KEY: misc_binding
; LOCATION: 7552..7576
; OTHER INFORMATION: 10-286-289.probe
; NAME/KEY: misc_binding
; LOCATION: 7607..7631
; OTHER INFORMATION: 10-286-345.probe
; NAME/KEY: misc_binding
; LOCATION: 7637..7661
; OTHER INFORMATION: 10-286-375.probe
; NAME/KEY: misc_binding
; LOCATION: 17246..17270
; OTHER INFORMATION: 12-425-57.probe
; NAME/KEY: misc_binding
; LOCATION: 21583..21607
; OTHER INFORMATION: 12-421-140.probe
; NAME/KEY: misc_binding
; LOCATION: 36959..36983
; OTHER INFORMATION: 10-523-232.probe
; NAME/KEY: misc_binding
; LOCATION: 45202..45226
; OTHER INFORMATION: 10-289-201.probe
; NAME/KEY: misc_binding
; LOCATION: 45729..45753
; OTHER INFORMATION: 10-290-37.probe
; NAME/KEY: misc_binding
; LOCATION: 46017..46041
; OTHER INFORMATION: 10-290-326.probe
; US-09-671-317-485

```

Query Match 18.4%; Score 27.8; DB 4; Length 49312;
 Best Local Similarity 65.1%; Pred. No. 16;

```
Matches 41; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 52 TTCAAGACATGGTGTGAAGCGCCGATCAACCTTCTGCGCTATTAAATGTCAT 111
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 23699 TTCAAGACGACGCTGGGCAACATAGCAAGAGCCCGCTCTCTATTAAAAAATATAA 23758
QY 112 TAA 114
    ||
Db 23759 AAA 23761

RESULT 14
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
```

```

; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

```

```

Query Match      18.1%; Score 27.4; DB 4; Length 1664976;
Best Local Similarity 55.9%; Pred. No. 91;
Matches 52; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 28 AATGTTGCCTTGGGGTTCCTGTTTCAACAGCATGGTGTGAAGCGCGCATCAACCTTC 87
Db 640752 AATGAAAGAAATTTGGGTAACCTCGTATACCATTTGTTGTCGATATTTCTATCTTA 640693

Qy 88 TCTGCTATTAAATAAAATGCTATAAATCAT 120
Db 640692 TTGCTAACATCAATTATAGTGATATAAATCAT 640660

```

```

RESULT 15
US-08-976-259-87
; Sequence 87, Application US/08976259
; Patent No. 6316609
; GENERAL INFORMATION:
; APPLICANT: Dillion, Patrick J.
; APPLICANT: Choi, Gil H.
; APPLICANT: Welch, Rodney A.
; TITLE OF INVENTION: Nucleotide Sequence of Escherichia coli
; Patent No. 6316609
; NUMBER OF SEQUENCES: 142
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Ave, N.W., Suite 600
; CITY: Washington
; STATE: DC

```

```

; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,259
; FILING DATE: Herewith
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/031,626 AND US 60/061,953
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/POCKET NUMBER: 1488.0740002/EKS/CSM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 87:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 595 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-976-259-87

Query Match      18.0%; Score 27.2; DB 4; Length 595;
Best Local Similarity 67.9%; Pred. No. 3.5;
Matches 38; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 95 ATTAATAAATAATGTCATAAATCATCTCTGCAAGTGGCAATTCCTCAAGATAT 150
Db 368 ATTAAGCAACATGCGAGATTCTCTCATCCACCGGCAACTTCTCTAATAT 423

Search completed: August 10, 2004, 21:58:15
Job time : 14.9751 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 10, 2004, 17:10:22 ; Search time 601.675 Seconds
(without alignments)
10877.634 Million cell updates/sec

Title: US-10-001-857-41

Perfect score: 151
Sequence: 1 ccgcgcggcaggtaacctaa.....gcaaatctctcaagaatatg 151

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.:

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inrv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	151	100.0	151	6	AX523453	AX523453 Sequence
C 2	140.4	93.0	3096	6	AX523454	AX523454 Sequence
C 3	106.4	70.5	160115	9	AL161447	AL161447 Human DNA
C 4	106.4	70.5	320902	2	AL161456	AL161456 Homo sapi
C 5	84.8	56.2	2488	6	AX537452	AX537452 Sequence
C 6	83.8	55.5	815	6	AX523523	AX523523 Sequence
C 7	83.2	55.1	1839	9	AK025266	AK025266 Homo sapi
C 8	72.6	48.1	281804	2	AC134869	AC134869 Mus muscu
C 9	70.6	46.8	253611	2	AC111473	AC111473 Rattus no
C 10	70.6	46.8	271861	2	AC108541	AC108541 Rattus no
C 11	67.4	44.6	2552	10	AY102701	AY102701 Mus muscu
C 12	67.4	44.6	2594	10	BC056435	BC056435 Mus muscu
C 13	63.2	41.9	2543	10	AF272892	AF272892 Rattus no
C 14	53	35.1	2467	5	BC053286	BC053286 Danio rer
C 15	53	35.1	147876	2	AC102010	AC102010 Mus muscu
C 16	36.6	24.2	3215	14	AB074755	AB074755 Hepatitis
C 17	35.8	23.7	70276	9	AC073095	AC073095 Homo sapi
C 18	35	23.2	266	14	AF276569	AF276569 Hepatitis
C 19	35	23.2	266	14	AF276570	AF276570 Hepatitis
C 20	35	23.2	266	14	AF276571	AF276571 Hepatitis
C 21	35	23.2	266	14	AF276572	AF276572 Hepatitis
C 22	35	23.2	266	14	AF276573	AF276573 Hepatitis
C 23	35	23.2	266	14	AF276574	AF276574 Hepatitis
C 24	35	23.2	266	14	AF276575	AF276575 Hepatitis
C 25	35	23.2	266	14	AF276576	AF276576 Hepatitis
C 26	35	23.2	465	14	AF378791	AF378791 Hepatitis
C 27	35	23.2	3202	14	AY217374	AY217374 Hepatitis
C 28	35	23.2	3206	14	AB112408	AB112408 Hepatitis
C 29	35	23.2	3215	14	AB112471	AB112471 Hepatitis
C 30	35	23.2	3215	14	AF068756	AF068756 Hepatitis
C 31	35	23.2	3215	14	AF223960	AF223960 Hepatitis
C 32	34.8	23.0	172542	2	AC146320	AC146320 Papio anu
C 33	34.8	23.0	181792	9	AC093126	AC093126 Papio anu
C 34	34.6	22.9	506	14	AF325702	AF325702 Hepatitis
C 35	34.6	22.9	512	14	AF325703	AF325703 Hepatitis
C 36	34.6	22.9	25400	3	CEP59A2	Z34801 Caenorhabdi
C 37	34.2	22.6	159994	2	AC073471	AC073471 Homo sapi
C 38	34	22.5	466	14	AY251136	AY251136 Hepatitis
C 39	33.8	22.4	5751	6	BD181579	BD181579 Novel gen
C 40	33.8	22.4	5751	9	AB058692	AB058692 Homo sapi
C 41	33.8	22.4	23631	9	HSV311G7	Z69304 Human DNA s
C 42	33.8	22.4	161078	9	AP003500	AP003500 Homo sapi
C 43	33.8	22.4	161286	2	AC025120	AC025120 Homo sapi
C 44	33.8	22.4	168467	10	AC123798	AC123798 Mus muscu
C 45	33.8	22.4	180303	9	AP003782	AP003782 Homo sapi

ALIGNMENTS

RESULT 1	AX523453	Sequence 41	151 bp	DNA	linear	PAT 24-OCT-2002
LOCUS	AX523453	Sequence 41	from Patent WO02064788.			
DEFINITION	AX523453					
ACCESSION	AX523453					
VERSION	AX523453.1	GI:24412349				
KEYWORDS						
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.					
REFERENCE	1					
AUTHORS	Macina,R.A., Recipon,H., Chen,S.Y., Sun,Y. and Liu,C.					
TITLE	Compositions and methods relating to lung specific genes and proteins					

```

JOURNAL Patent: WO 02064788-A 41 22-AUG-2002;
FEATURES Diadexus, Inc. (US)
SOURCE Location/Qualifiers
1..151
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 100.0%; Score 151; DB 6; Length 151;
Best Local Similarity 100.0%; Pred. No. 7.6e-38;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGCCCGGCGAGGTACCTAAACAGGCCAAATGTTGCCCTTGGGGTTCCTGTTTCAACAGC 60
Db 1 CCGCCCGGCGAGGTACCTAAACAGGCCAAATGTTGCCCTTGGGGTTCCTGTTTCAACAGC 60

Qy 61 ATGTTGTAAGCGCGCGCATCAACCTTCTCGCTATTAAATAAAATGTCATAAACTCAT 120
Db 61 ATGTTGTAAGCGCGCGCATCAACCTTCTCGCTATTAAATAAAATGTCATAAACTCAT 120

Qy 121 CCTGCAGGTGCAAAATCTCTCAAGAAATAG 151
Db 121 CCTGCAGGTGCAAAATCTCTCAAGAAATAG 151

RESULT 2
LOCUS AX523454 3096 bp DNA linear PAT 24-OCT-2002
DEFINITION Sequence 42 from Patent WO02064788.
ACCESSION AX523454
VERSION AX523454.1 GI:24412350
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1
AUTHORS Macina,R.A., Recipon,H., Chen,S.Y., Sun,Y. and Liu,C.
TITLE Compositions and methods relating to lung specific genes and
proteins
JOURNAL Patent: WO 02064788-A 42 22-AUG-2002;
Diadexus, Inc. (US)
FEATURES Location/Qualifiers
source 1..3096
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 93.0%; Score 140.4; DB 6; Length 3096;
Best Local Similarity 99.3%; Pred. No. 2.9e-34;
Matches 141; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 CAGTACTTAACAGGCCAAATGTTGCCCTTGGGGTTCCTGTTTCAACAGCATGTTGTA 69
Db 2052 CAGTACTTAACAGGCCAAATGTTGCCCTTGGGGTTCCTGTTTCAACAGCATGTTGTA 1993

Qy 70 AGCGCCGATCAACCTTCTCGCTATTAAATAAAATGTCATAAACTCATCTGCAAGG 129
Db 1992 AGCGCTCATCAACCTTCTCGCTATTAAATAAAATGTCATAAACTCATCTGCAAGG 1933

Qy 130 TGGCAAAATCTCTCAAGAAATAG 151
Db 1932 TGGCAAAATCTCTCAAGAAATAG 1911

RESULT 3
AL161447/c
LOCUS AL161447 160115 bp DNA linear PRI 16-MAR-2002
DEFINITION Human DNA sequence from clone RP11-379P1 on chromosome 9, complete
sequence.
ACCESSION AL161447

```

VERSION
KEYWORDS
SOURCE
ORGANISM

AL161447.20 GI:19572746

HTG.
Homo sapiens (human)

REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (16-MAR-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Mar 21, 2002 this sequence version replaced gi:15787718. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

COMMENT

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em', EMBL; SW', SWISSPROT; Tr', TREMBL; Wp', WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9> RP11-379P1 is from the library RP11-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECT08: pBACE3.6 This sequence is the entire insert of clone RP11-379P1 The true left end of clone RP11-280P22 is at 137145 in this sequence. The true right end of clone RP11-65G15 is at 38497 in this sequence.

FEATURES

Location/Qualifiers
source 1..160115
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-379P1"
/clone_lib="RP11-11.2"
complement(32505..32575)
/note="Sequence from uni-directional dGTP big dye terminator reads only."
74372..74537
/note="Single clone region. Assembly confirmed by restriction digest data."
74706..74920
/note="Single clone region. Assembly confirmed by restriction digest data."
complement(74706..74920)
/note="Sequence from uni-directional dGTP big dye terminator reads only."
74921
/note="Tandem repeat. Forced join. Gap size estimated to be approximately 200bp by restriction digest data."
123108
/note="Tandem repeat. Forced join. Gap size estimated to be approximately 300bp by restriction digest data."

ORIGIN

Query Match 70.5%; Score 106.4; DB 9; Length 160115;
Best Local Similarity 99.1%; Pred. No. 4.6e-23;
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Oy 10 CAGGTACCTAAACAGCCAAATGTTCCCTTGGGTTCTCTGTTTAAACAGCATGGTGTGA 69
Db 35038 CAGGTACCTAAACAGCCAAATGTTCCCTTGGGTTCTCTGTTTAAACAGCATGGTGTGA 34979
Oy 70 AGCGCGCATCAACCTTCTCTGCTATTAAATAAAATGTCATAAACT 117
Db 34978 AGCGCGCATCAACCTTCTCTGCTATTAAATAAAATGTCATAAACT 34931

RESULT 4
LOCUS AL161456/c 320902 bp DNA linear HTG 15-JAN-2002
DEFINITION Homo sapiens chromosome 9 clone RP11-85C21, 85 unordered pieces.
ACCESSION AL161456
VERSION AL161456.16 GI:15485122
KEYWORDS HTG; HTGS PHASE1; HTGS_CANCELLED.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 Plumb,B.
Direct Submission
Submitted (14-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Sep 6, 2001 this sequence version replaced gi:11322794.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bA85C21
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M77815; 1% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 289121 bases at least Q40
Consensus quality: 300401 bases at least Q30
Consensus quality: 306412 bases at least Q20
Insert size: 312502; sum-of-contigs
Insert size: 179280; 3.8% error; agarose-fp
Quality coverage: 3.24x in Q20 bases; sum-of-contigs Quality
coverage: 8.83x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 85 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 2435: contig of 2435 bp in length
* 2436 2535: gap of 100 bp
* 2536 9914: contig of 7379 bp in length
* 9915 10014: gap of 100 bp
* 10015 14224: contig of 4210 bp in length
* 14225 14324: gap of 100 bp
* 14325 18383: contig of 4059 bp in length
* 18384 21333: gap of 100 bp
* 21333 21433: contig of 2850 bp in length
* 21434 24294: contig of 2861 bp in length
* 24295 24394: gap of 100 bp
* 24395 28746: contig of 4352 bp in length
* 28747 28846: gap of 100 bp
* 28847 34452: contig of 5606 bp in length
* 34453 34553: gap of 100 bp
* 34553 36649: contig of 2097 bp in length

36750 36749: gap of 100 bp
36750 39860: contig of 3211 bp in length
39861 40060: gap of 100 bp
40061 42135: contig of 2075 bp in length
42136 42235: gap of 100 bp
42236 45658: contig of 3423 bp in length
45659 45758: gap of 100 bp
45759 51237: contig of 5479 bp in length
51238 51337: gap of 100 bp
51338 56833: contig of 5496 bp in length
56834 61915: contig of 4982 bp in length
61916 62015: gap of 100 bp
62016 64599: contig of 2584 bp in length
64600 66720: contig of 2021 bp in length
66721 66820: gap of 100 bp
66821 69328: contig of 2508 bp in length
69329 69428: gap of 100 bp
69429 72905: contig of 3477 bp in length
72906 73005: gap of 100 bp
73006 77059: contig of 4054 bp in length
77060 79159: gap of 100 bp
79160 79297: contig of 2038 bp in length
79298 84185: contig of 4888 bp in length
84186 84285: gap of 100 bp
84286 88969: contig of 4684 bp in length
88970 89069: gap of 100 bp
89070 92199: contig of 3130 bp in length
92200 92299: gap of 100 bp
92300 94921: contig of 2622 bp in length
94922 95021: gap of 100 bp
95022 98599: contig of 3578 bp in length
98600 107281: contig of 8582 bp in length
107282 107381: gap of 100 bp
107382 112698: contig of 5317 bp in length
112699 112798: gap of 100 bp
112799 114885: contig of 2087 bp in length
114886 114985: gap of 100 bp
114986 117732: contig of 2747 bp in length
117733 117832: gap of 100 bp
117833 120717: contig of 2885 bp in length
120718 120817: gap of 100 bp
120818 123116: contig of 2299 bp in length
123117 126117: contig of 2901 bp in length
126118 126217: gap of 100 bp
126218 129518: contig of 3301 bp in length
129519 129618: gap of 100 bp
129619 134096: contig of 4478 bp in length
134097 134196: gap of 100 bp
134197 139337: contig of 5141 bp in length
139338 139437: gap of 100 bp
139438 141501: contig of 2064 bp in length
141502 141601: gap of 100 bp
141602 143685: contig of 2084 bp in length
143686 143785: gap of 100 bp
143786 146147: contig of 2362 bp in length
146148 146247: gap of 100 bp
146248 149763: contig of 3516 bp in length
149764 149863: gap of 100 bp
149864 152002: contig of 2139 bp in length
152003 152102: gap of 100 bp
152103 155247: contig of 3145 bp in length
155248 155347: gap of 100 bp
155348 157398: contig of 2051 bp in length
157399 157498: gap of 100 bp
157499 159799: contig of 2301 bp in length
159800 159899: gap of 100 bp
159900 163397: contig of 3498 bp in length
163398 163497: gap of 100 bp

```

* 163498 166241: contig of 2744 bp in length
* 166242 166341: gap of 100 bp
* 166342 168447: contig of 2106 bp in length
* 168447 168547: gap of 100 bp
* 168547 171536: contig of 2989 bp in length
* 171536 171636: gap of 100 bp
* 171636 175769: contig of 4133 bp in length
* 175769 175869: gap of 100 bp
* 175869 183977: contig of 8108 bp in length
* 183977 184077: gap of 100 bp
* 184077 186608: contig of 2531 bp in length
* 186608 186708: gap of 100 bp
* 186708 190121: contig of 3413 bp in length
* 190121 190221: gap of 100 bp
* 190221 197833: contig of 7612 bp in length
* 197833 197933: gap of 100 bp
* 197933 200266: contig of 2333 bp in length
* 200266 200366: gap of 100 bp
* 200366 203220: contig of 2854 bp in length
* 203220 203320: gap of 100 bp
* 203320 208956: contig of 5636 bp in length
* 208956 209057: gap of 100 bp
* 209057 211798: contig of 2742 bp in length
* 211798 211898: gap of 100 bp
* 211898 214088: contig of 2190 bp in length
* 214088 214188: gap of 100 bp
* 214188 217324: contig of 3136 bp in length
* 217324 217424: gap of 100 bp
* 217424 220661: contig of 3237 bp in length
* 220661 225678: contig of 4917 bp in length
* 225678 225779: gap of 100 bp
* 225779 228798: contig of 3020 bp in length
* 228798 231289: gap of 100 bp
* 231289 231389: contig of 2391 bp in length
* 231389 235056: contig of 3667 bp in length
* 235056 235156: gap of 100 bp
* 235156 238252: contig of 3096 bp in length
* 238252 238352: gap of 100 bp
* 238352 241060: contig of 2708 bp in length
* 241060 241160: gap of 100 bp
* 241160 244514: contig of 3354 bp in length
* 244514 247447: contig of 2833 bp in length
* 247447 247547: gap of 100 bp
* 247547 252282: contig of 4735 bp in length
* 252282 252382: gap of 100 bp
* 252382 254818: contig of 2436 bp in length
* 254818 254919: gap of 100 bp
* 254919 257094: contig of 2176 bp in length
* 257094 257194: gap of 100 bp
* 257194 259885: contig of 2691 bp in length
* 259885 262579: gap of 100 bp
* 262579 262679: contig of 2594 bp in length
* 262679 267289: contig of 4610 bp in length
* 267289 267389: gap of 100 bp
* 267389 271789: contig of 4400 bp in length
* 271789 271889: gap of 100 bp
* 271889 280376: contig of 8487 bp in length
* 280376 280476: gap of 100 bp
* 280476 286094: contig of 5618 bp in length
* 286094 286194: gap of 100 bp
* 286194 293197: contig of 7003 bp in length
* 293197 293297: gap of 100 bp
* 293297 293398: gap of 100 bp

Query Match 70.5%; Score 106.4; DB 2; Length 320902;
Best Local Similarity 99.1%; Pred. No. 5e-23;
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 CAGGTACCTAACAGCCAAATGTCCTTTGGGGTTCCTGTTTCAACAGCATGCTGTA 69
|||||

Db 261433 CAGGTACCTAACAGCCAAATGTCCTTTGGGGTTCCTGTTTCAACAGCATGCTGTA 261374
Qy 70 AGCGCCGATCAACCTTCTGCTTAAATAAAATGTCATAAAT 117
|||||
Db 261373 AGCGTGCATCAACCTTCTGCTTAAATAAAATGTCATAAAT 261326
|||||
RESULT 5
AX537452/c 2488 bp DNA linear PAT 23-NOV-2002
LOCUS AX537452
DEFINITION Sequence 23 from Patent WO02070709.
ACCESSION AX537452
VERSION AX537452.1 GI:25269238
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Lai, P.G., Baughn, M.R., Yao, M.G., Wallia, N.K., Elliot, V.S., Xu, Y.,
Honchell, C.D., Yue, H., Ding, L., Gietzen, K.J., Ison, C.H., Lu, D.A.,
Hafalia, A.J., Ghandi, A.R., Thangavelu, K., Sanjanwala, M.M.,
Tang, Y.T., Ramkumar, J., Griffin, J.A., Swarnaker, A., Azimzai, Y.,
Sapperstein, S.K., Burford, N., Lee, E.A., Lu, Y., Tran, U.K. and
Marquis, J.P.
TITLE Molecules for disease detection and treatment
JOURNAL Patent: WO 02070709-A 23 12-SEP-2002;
Incyte Genomics, Inc. (US)
FEATURES
source 1. .2488
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 7111920CB1"

ORIGIN

Query Match 56.2%; Score 84.8; DB 6; Length 2488;
Best Local Similarity 88.5%; Pred. No. 2.6e-16;
Matches 92; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
Qy 10 CAGGTACCTAACAGCCAAATGTCCTTTGGGGTTCCTGTTTCAACAGCATGCTGTA 69
|||||
Db 1554 CAGGTACCTAACAGCCAAATGTCCTTTGGGGTTCCTGTTTCAACAGCATGCTGTA 1495
|||||
Qy 70 AGCGCCGATCAACCTTCTGCTTAAATAAAATGTCATA 113
|||||
Db 1494 AGCGTGCATCAACCTTCTGCTTAAATAAAATGTCATA 1451
|||||

RESULT 6

AX523523 815 bp DNA linear PAT 24-OCT-2002
LOCUS AX523523
DEFINITION Sequence 111 from Patent WO02064788.
ACCESSION AX523523
VERSION AX523523.1 GI:24412419
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Macina, R.A., Recipon, H., Chen, S.Y., Sun, Y. and Liu, C.
TITLE Compositions and methods relating to lung specific genes and
proteins
JOURNAL Patent: WO 02064788-A 111 22-AUG-2002;
Diadexus, Inc. (US)
FEATURES
source 1. .815
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

* 2326 2425: gap of unknown length
 * 2426 4173: contig of 1748 bp in length
 * 4174 473: gap of unknown length
 * 4274 6330: contig of 2857 bp in length
 * 7031 7030: gap of unknown length
 * 10444 10444: contig of 3414 bp in length
 * 10545 14581: contig of 4037 bp in length
 * 14582 14681: gap of unknown length
 * 14682 17225: contig of 2544 bp in length
 * 17226 17325: gap of unknown length
 * 17326 22780: contig of 5455 bp in length
 * 22781 22880: gap of unknown length
 * 22881 34746: contig of 11866 bp in length
 * 34747 34846: gap of unknown length
 * 34847 51109: contig of 18263 bp in length
 * 51110 5209: gap of unknown length
 * 5210 65908: contig of 15699 bp in length
 * 65909 67008: gap of unknown length
 * 67009 85189: contig of 18181 bp in length
 * 85190 85289: gap of unknown length
 * 85290 119762: contig of 34473 bp in length
 * 119763 119862: gap of unknown length
 * 119863 190705: contig of 70843 bp in length
 * 190706 190805: gap of unknown length
 * 190806 281804: contig of 90999 bp in length.

FEATURES

source

1. 281804
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="UNK"
 /clone="RP23-137C6"
 1. 1074
 /note="assembly_name:Contig10"
 1175. 2325
 /note="assembly_name:Contig12"
 2426. 4173
 /note="assembly_name:Contig27"
 4274. 6930
 /note="assembly_name:Contig30"
 7031. 10444
 /note="assembly_name:Contig31"
 10545. 14581
 /note="assembly_name:Contig32"
 14682. 17225
 /note="assembly_name:Contig33"
 17326. 22780
 /note="assembly_name:Contig34"
 22881. 34746
 /note="assembly_name:Contig35"
 34847. 51109
 /note="assembly_name:Contig36"
 51210. 65908
 /note="assembly_name:Contig37"
 67009. 85189
 /note="assembly_name:Contig38"
 85290. 119762
 /note="assembly_name:Contig39"
 119863. 190705
 /note="assembly_name:Contig40"
 190806. 281804
 /note="assembly_name:Contig41"

ORIGIN

Query Match 48.1%; Score 72.6; DB 2; Length 281804;
 Best Local Similarity 85.3%; Pred. No. 4.1e-12;
 Matches 81; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
 10 CAGGTACTAAACAGGCCAATGTGCTTTGGGGTTCCCTGTTTCAACAGCATGGTGA 69
 |||||
 Db 241454 CAGGTCTTAGCAGGCCAGATGTGTCTGAGGCTTCCTGTTTCAACACATAGTGA 241395
 |||||

QY 70 AGCGCGCATCAACCTCTCTGCGCTATTAAATAA 104
 |||||
 Db 241394 AGCGCTGCATCAACCTCTCTGCGCTGCAAAACAA 241360
 |||||

RESULT 9

AC111473/C

LOCUS AC111473 253611 bp DNA linear HTG 13-MAY-2003
 Rattus norvegicus clone CH230-82L20, WORKING DRAFT SEQUENCE, 3
 unordered pieces.

ACCESSION

AC111473

HTG, HTGS, PHASE1, HTGS, DRAFT; HTGS_FULLTOP.

KEYWORDS

Rattus norvegicus (Norway rat)

SOURCE

Rattus norvegicus

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 253611)
 Muzny D, Marie, Metzker, M, Lee, Abranzone, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hughes, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, B., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensuhewa, L., Loulseghe, H., Lozano, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwankwelen, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Pioppor, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
 Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhauser, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G. and Gibbs, R.A.

Direct Submission

Unpublished

2 (bases 1 to 253611)

Worley, K.C.

Direct Submission

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL Submitted (19-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 253611)

AUTHORS Rat Genome Sequencing Consortium.

TITLE Direct Submission

JOURNAL Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On May 13, 2003 this sequence version replaced gi:24942245. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GKMA

Center clone name: CH230-82L20

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 242469 bases at least Q40

Consensus quality: 245181 bases at least Q30

Consensus quality: 246854 bases at least Q20

Estimated insert size: 252908; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)

* NOTE: This sequence may represent more than one clone.

* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

* 1 250191: contig of 250191 bp in length

* 250192 250291: gap of unknown length

* 250292 252060: contig of 1769 bp in length

* 252061 252160: gap of unknown length

* 252161 253611: contig of 1451 bp in length.

FEATURES Location/Qualifiers

source 1..253611

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-82L20"

1..1668

/note="wgs_end_extension clone_end:T7"

3547..4345

/note="clone boundary clone_end:T7"

site:EcORI

end_sequence:BH268278"

complement(248438..249256)

/note="clone boundary clone_end:Sp6"

site:EcORI

end_sequence:BH268279"

ORIGIN

Query Match 46.8%; Score 70.6; DB 2; Length 253611; Best Local Similarity 72.8%; Pred. No. 1.8e-11; Matches 91; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 10 CAGGTACCTAAACAGCGCAAAATGTCCTTGGGTTCTCTGTTCAACAGCATGGTGGA 69

DB 26933 CAAAGTCTTAAACAGCGCAGATGCTGCTGAGGCTCTGCTTCAACACATAGTGTG 26874

QY 70 AGCGCGCATCAACCTTCTCTGCTATTAAAAATAAAATGTCATAAACTCATCTCGCAAG 129

DB 26873 AGCGTGCATGCACCTTCTCTGCTGTCAAATAATTGTTACATPACATCAATTATGT 26814

QY 130 TGGCA 134

DB 26813 AGACA 26809

RESULT 10

AC108541

LOCUS

DEFINITION Rattus norvegicus clone CH230-137C13, WORKING DRAFT SEQUENCE, 7

AC108541 271861 bp DNA linear HTG 13-MAY-2003

unordered pieces.

ACCESSION AC108541

VERSION AC108541.5 GI:30580127

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 271861)

Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Diya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Faves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,X., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,D., Lorenschwah,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puaro,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reich,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,

Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 271861)
Worley, K.C.

Direct Submission
Submitted (31-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 271861)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 13, 2003 this sequence version replaced gi:23123542. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GJBM
Center clone name: CH230-137C13
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 233080 bases at least Q40
Consensus quality: 237201 bases at least Q30
Consensus quality: 240317 bases at least Q20
Estimated insert size: 241904; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 4170: contig of 4170 bp in length
* 4171 4270: gap of unknown length
* 4271 175611: contig of 171341 bp in length
* 175612 175711: gap of unknown length
* 175712 182155: contig of 6444 bp in length
* 182156 182255: gap of unknown length
* 182256 265469: contig of 83214 bp in length
* 265470 265569: gap of unknown length
* 265570 267249: contig of 1680 bp in length
* 267250 267250: gap of unknown length
* 267250 268529: contig of 1180 bp in length

* 268530 268629: gap of unknown length
* 268630 271861: contig of 3232 bp in length.

FEATURES
source
1..271861
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-137C13"
1..1466
/note="wgs end extension
clone_end:Sp6"
complement(3254..3667)
/note="clone boundary
clone_end:Sp6
site:EcoRI
end_sequence:BH355884"
4271..5454
/note="wgs contig"
21028..22738
/note="wgs contig"
175712..176921
/note="wgs contig"
182256..183591
/note="wgs contig"

misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

ORIGIN
Query Match 46.8%; Score 70.6; DB 2; Length 271861;
Best Local Similarity 72.8%; Pred. No. 1.8e-11;
Matches 91; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 10 CAGGTACTTAAACAGGCAAAATGCTTGGGGTTCCTTTCAACAGCATGTGTGA 69
|||
Db 97113 CAAGTTCTTAAACAGGCAAGANGCTGCTGAGCTCTCTCTCAACAATAGTGTGG 97172
|||

QY 70 AGCCCGCATCAACCTTCTCTGCTATTAAATAAATGTCAAACTCATCTCGCAAGG 129
|||
Db 97173 AGCGCTGCATCGACCTTCTGCTGCTGCAAAATAATTGTACATTACAATTAATGT 97232
|||

QY 130 TGGCA 134
|||
Db 97233 AGACA 97237

RESULT 11
AY102701/c
LOCUS AY102701 Mus musculus embryonic growth-associated protein EGAP mRNA, complete cds. 2552 bp mRNA linear ROD 23-JUN-2002
ACCESSION AY102701
VERSION AY102701.1 GI:21539895
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2552)
AUTHORS Wenzlau, J.M. and Weiser-Evans, M.C.M.
TITLE Mouse Embryonic Growth Associated Protein (EGAP)
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2552)
AUTHORS Wenzlau, J.M. and Weiser-Evans, M.C.M.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2002) Pediatrics, University of Colorado HSC, 4200 East Ninth Ave, Denver, CO 80262, USA

FEATURES
source
1..2552
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/chromosome="11"
/dev_stage="embryo"
55..2232
/codon_start=1

CDS


```

Db      1513 CAGGTTCTTAACGAGCGCAGATGTGTCTCTGAGGCTCCTGTGTTTCAACAACATAGTATGA 1454
QY      70 AGCGCGCATCAACCTTCTGCGCT 94
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1453 AGCGTGCATCAACCTTCTGCGCT 1429

RESULT 13
LOCUS   AF272892/c
DEFINITION Rattus norvegicus corneal wound healing related protein mRNA, ROD 15-SEP-2000 complete cds.
ACCESSION AF272892
VERSION   AF272892.1 GI:8926319
KEYWORDS
SOURCE    Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 2543)
AUTHORS   Yi, X.J., Li, X.F. and Yu, F.S.
TITLE      A novel epithelial wound-related gene is abundantly expressed in developing rat cornea and skin
JOURNAL    Curr. Eye Res. 20 (5), 430-440 (2000)
MEDLINE    20314365
PUBMED     10855038
REFERENCE 2 (bases 1 to 2543)
AUTHORS   Li, X. and Yu, F.
TITLE      Direct Submission
JOURNAL    Submitted (26-MAY-2000) The Schepens Eye Research, 20 Staniford Street, Boston, MA 02114, USA
FEATURES   Location/Qualifiers
            source
              1..2543
              /organism="Rattus norvegicus"
              /mol_type="mRNA"
              /db_xref="taxon:10116"
              64..2241
              /codon_start=1
              /product="corneal wound healing related protein"
              /protein_id="AA081791.1"
              /db_xref="GI:8926320"
              /translation="MVNKAADVDDASGWLNVPEKMEKSKTSWVDITQDFEDACRELK
              LGELHDKLFLGLFAMSAIEMMDPKMDAGMIGNOVNRKVLNFEQAVKQGTIKIKDLSL
              PELIGIMDTFCCLITWLGHSIAQVFTCLYIHPDPFIEDPAKFAIGLTKICDIA
              REKNKAAAFEEEDFQSWTYGFKMANGVTDLRVTGMLKVEDDMQRVKSTRSQGEE
              RDEVELEHQCLAAFSRVFTRVLLVLIATFKETSAVAAGKQVMQVADLLSLAH
              TSLHGIQAGNTTGKGDHPIMMGPELVNQRLLPPTPPYAKIKREEMVNYFSRLID
              RIKTVCVNLNPLHCLIDFFCFSESPCVLSRLQTTFLVDNKKVFGTHLMQDMV
              KDAPRSFVSPVLSPKCLYNNHQKDCIDSFVTHCVRFCSLVQIHGHNRARQDKL
              GHILEFATQDAEAKVDAAHLTMLKQEPQROHACLGITWLVHSLRIMIOYLLSGF
              DLELYSMHYEYIYVWYLSFLYAWLMSLTSRAGSOMAEERIMEEQKGRSSKTKKK
              KVRPLSREITMSQAYQNCAGMFKTVAEDMGKVRKPFELDSEQVREYHRRFAPFN
              SVMTPPPHVLQFKEMSDLSKYPPOPPELTVAAASHFOQAKMILESIPNADREVS
              ILKVAKPNFVMKLLAGGHKKESKVPPEPDSFVHKYPPVVKLV"

ORIGIN
Query Match 41.98; Score 63.2; DB 10; Length 2543;
Best Local Similarity 74.18; Pred. No. 2.5e-09;
Matches 80; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY      10 CAGGTACCTAAACAGGCAAAATGTGCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGA 69
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1535 CAAGTTCCTAAACAGGCGAGATGCTGCTGAGGCTCCTGCTTCAACAACATAGTGTGG 1476
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY      70 AGCGCGCATCAACCTTCTGCGCTATATAAATGAATGTCATAACT 117
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1475 AGCGTGCATGACCTTCTGCTGCTATCTTGCAAGGTGCAAACTCT 1428
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
LOCUS   BC053286/c
DEFINITION Danio rerio cDNA clone MGC:64157 IMAGE:6797204, complete cds.
ACCESSION BC053286
VERSION   BC053286.1 GI:31418975
KEYWORDS
SOURCE    Danio rerio (zebrafish)
ORGANISM Danio rerio
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 2467)
AUTHORS   Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J.J., Hsieh, F.K., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
TITLE      Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
FEATURES   Location/Qualifiers
            source
              1..2467
              Strausberg, R.
              Direct Submission
              Submitted (02-JUN-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
              NIH-MGC Project URL: http://mgc.nci.nih.gov
              Contact: MGC help desk
              Email: qcaps-xemail.nih.gov
              Tissue Procurement: Leonard I. Zon, M.D.
              cDNA Library Preparation: Invitrogen Corp
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
              DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
              Web site: http://www-shgc.stanford.edu
              Contact: (Dickson, Mark) mcd@paxil.stanford.edu
              Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
Series: IRAC Plate: 117 Row: n Column: 11
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.
FEATURES   Location/Qualifiers
            source
              1..2467
              /organism="Danio rerio"
              /mol_type="mRNA"
              /db_xref="taxon:7955"
              /clone="MGC:64157 IMAGE:6797204"
              /tissue_type="Kidney, zebrafish"
              /clone_lib="NCI CGAP_ZKId1"
              /lab_host="DH10B"
              /note="Vector: pCMV-SPORT6.1"
              58..2232
              /codon_start=1
              /product="Unknown (protein for MGC:64157)"
              /protein_id="AAH53286.1"

```

```

/db_xref="GI:31418976"
/translation="MWKSSVEEEGGWGLIPEKMRNNANWVDVTQEPKGACKELK
GELLHDKPLGLEFMSAIGMDPMQDAGMIGQVNRKVLNFOQAVDEAIRVKDLSIP
ELIGMDTCFCCLITWEGHSLAQFTCLYVHNPDLIQDPALKAFLGKLCIDTAR
EKVNAVFEEDFOAMTYFKMANNVTDLRVTGMLKQVDELDQKRVKSTRQGBQR
DPEVLDHQOCLALFSRVAFTRLLSALISFTKKTSAVSEACKLMSQAADLLPAVHA
TIQYGLQSDNTTKGDPHIMMGFEPLVNGRLPPTPRYAKIIKREMYNYSKLLIER
IKSVCEVINITNHSILDFCFSESPCVLSRLQTTFLDINKKVGTHLMQDMIK
DALRCFVSPVLSKSNANNHQADYIDSFVTHCTRFPSLQIHGHNRARQDKIG
HILEEPATQDEAKRYDAALHGLMKLEPQRQHLACLGITWILYHNIRIMQYLLSGFE
LELYSMHEYYIYVWSEFLYAWLMTSLGRADSOAEERILEEOLKVRSSKKKKK
KARPLSKTMSQAYONMCAQGYKWTIALDMORVKRPOPELSEOVRYEHRPAPNS
VVTTPPVHVIQPKEMSDLKXNPPPSADFLYMAASKHFEQAKLLLENVSPDAEVNRI
LKVAKENIVIMKLLAGGHKTKALPELDFSAHKYIPPIVKIL"
misc_feature
139..693
/notes="Mak10; Region: Mak10 subunit, NatC
N(alpha)-terminal acetyltransferase. NatC
N(alpha)-terminal acetyltransferase contains Mak10p,
Mak31p and Mak3p subunits. All three subunits are
associated with each other to form the active complex"
/db_xref="CDD:pfam04112"

```

ORIGIN

```

Query Match      35.1%; Score 53; DB 5; Length 2467;
Best Local Similarity 76.5%; Pred No; 4.8e-06;
Matches 65; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 10 CAGGTACCTAAACAGGCCAAATGTGCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGA 69
|||||
Db 1526 CAGGTGCCACACAGGCTAAATGCTGCGTGGCGTCCAGCTTCATCAGCAGACCGTGA 1467
|||||

QY 70 AGCGCGCATCAACCTCTCTGCGCT 94
|||
Db 1466 AGAGCAGCGTCCACCTCTCTGCGCT 1442

```

RESULT 15

```

AC102010/c
LOCUS AC102010 147876 bp DNA linear HTG 21-MAR-2003
DEFINITION Mus musculus clone RP24-531M24, WORKING DRAFT SEQUENCE, 7 unordered
pieces.
ACCESSION AC102010
VERSION AC102010.3 GI:29135693
KEYWORDS HTG; HTGS_PHRASE1; HTGS_DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 147876)
Mus musculus, clone RP24-531M24
Unpublished
2 (bases 1 to 147876)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepey,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,

```

```

Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 147876)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepey,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
MacDonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 21, 2003 this sequence version replaced gi:28412092.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seg.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L17873
Center clone name: 531_M 24
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 146099 bases at least Q40
Consensus quality: 146869 bases at least Q30
Consensus quality: 147004 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 147276; sum-of-contigs
Quality coverage: 9.9 in Q20 bases; agarose-fp
Quality coverage: 10.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1 84958: contig of 84958 bp in length
* 84959 85058: gap of 100 bp
* 85059 87411: contig of 2353 bp in length
* 87412 87511: gap of 100 bp
* 87512 89933: contig of 2422 bp in length
* 89934 90033: gap of 100 bp
* 90034 94076: contig of 4043 bp in length
* 94077 94176: gap of 100 bp
* 94177 102536: contig of 8360 bp in length
* 102537 102636: gap of 100 bp
* 102637 118954: contig of 16318 bp in length
* 118955 119054: gap of 100 bp

```

TITLE

JOURNAL

COMMENT

FEATURES * 119055 147876: contig of 28822 bp in length.

Location/Qualifiers
1..147876
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-531M24"
/clone_lib="RPCI-24 Male Mouse BAC"
misc_feature 1..84958
/note="assembly_fragment"
clone_end:SP6
vector_side:left
85059..87411
/note="assembly_fragment"
misc_feature 87512..89933
/note="assembly_fragment"
90034..94076
/note="assembly_fragment"
misc_feature 94177..102536
/note="assembly_fragment"
misc_feature 102637..118954
/note="assembly_fragment"
119055..147876
/note="assembly_fragment"
clone_end:T7
vector_side:right"

ORIGIN

Query Match 35.1%; Score 53; DB 2; Length 147876;
Best Local Similarity 73.1%; Pred. No. 8.1e-06;
Matches 68; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
Qy 10 CAGGTACCTTAACAGCCCAATGTCCTTTGGGTTCTCTTCAACAGCATGGTGCA 69
Db 72729 CAGATTCTTAACAGGTAGATGATCTCTCAGGCTCTCTTCAACAGCATGGTGCA 72670
Qy 70 AGCGCCGCATCAACCTTCTCTGCCCTATTAAAAAT 102
Db 72669 AGCACTGCATCAAAAGTTCTGTGCAAGTGACT 72637

Search completed: August 11, 2004, 01:34:02
Job time : 605.675 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 10, 2004, 17:10:23 ; Search time 410.913 Seconds
(without alignments)
10973.574 Million cell updates/sec

Title: US-10-001-857-41

Perfect score: 151

Sequence: 1 ccgccgggcaggtagctacaa.....gcaattctctcaagaatatg 151

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estlin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hcc:*

9: gb_estl:*

10: gb_est2:*

11: gb_hcc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104.8	69.4	701	29	AG180176 Pan trogl
C 2	86.8	57.5	3071	11	AK050805 Mus muscu
C 3	75.2	49.8	812	14	CD654321 AGENCOURT
C 4	73.8	48.9	472	14	CF762228 CES000418

C 5	73.8	48.9	486	12	BI537770	BI537770 427999 MA
C 6	73.8	48.9	569	12	BI682349	BI682349 463408 MA
C 7	73.8	48.9	814	14	CB169246	CB169246 VBB603020
C 8	72.6	48.1	630	28	AZ870091	AZ870091 2M0182N22
C 9	68.4	45.3	974	10	BF179099	BF179099 601808136
C 10	67.4	44.6	409	9	AI020180	AI020180 ub15b06.r
C 11	67.4	44.6	426	10	BB727790	BB727790 B3727790
C 12	67.4	44.6	435	10	BB827250	BB827250 BB827250
C 13	67.4	44.6	446	10	BB832780	BB832780 BB832780
C 14	67.4	44.6	478	10	BE308753	BE308753 601091825
C 15	67.4	44.6	512	14	W54490	W54490 md09d10.r1
C 16	67.4	44.6	514	10	BF465930	BF465930 UI-M-CG0P
C 17	67.4	44.6	528	10	BE692994	BE692994 SAC448 MO
C 18	67.4	44.6	528	10	BE693019	BE693019 SAC473 MO
C 19	67.4	44.6	578	14	CA533399	CA533399 C0348H03-
C 20	67.4	44.6	618	14	CF175090	CF175090 B0948F03-
C 21	67.4	44.6	763	13	BU613022	BU613022 UI-M-FR0-
C 22	67.4	44.6	873	9	AU081000	AU081000 AU081000
C 23	67.4	44.6	900	14	CA977646	CA977646 AGENCOURT
C 24	66.4	44.0	595	10	BF730939	BF730939 mab79a08.
C 25	66.4	44.0	600	13	BU920897	BU920897 6071-74 M
C 26	65.8	43.6	434	10	BB811746	BB811746 BB811746
C 27	63.2	41.9	565	12	BM387958	BM387958 UI-R-CN1-
C 28	63.2	41.9	708	14	CF113422	CF113422 Shultzomi
C 29	62	41.1	416	10	BB820704	BB820704 BB820704
C 30	61.6	40.8	437	10	BF522938	BF522938 UI-R-Y0-a
C 31	61.6	40.8	482	14	CA792702	CA792702 AGENCOURT
C 32	61.6	40.8	603	9	AL782358	AL782358 AL782358
C 33	61.6	40.8	868	14	CF238691	CF238691 AGENCOURT
C 34	61.4	40.7	518	13	BY476464	BY476464 BY476464
C 35	57.8	38.3	300	10	BB192400	BB192400 BB192400
C 36	57.8	38.3	784	13	BU260643	BU260643 603503268
C 37	57.8	38.3	833	13	BU464042	BU464042 603773152
C 38	56.4	37.4	399	14	CF727314	CF727314 UI-M-HB0-
C 39	55.6	36.8	236	10	BF172850	BF172850 MYE0095 M
C 40	55	36.4	797	13	BX870686	BX870686 BX870686
C 41	54.6	36.2	303	10	BB273684	BB273684 BB273684
C 42	53.8	35.6	288	10	BB099049	BB099049 BB099049
C 43	53.8	35.6	456	10	BB730820	BB730820 BB730820
C 44	50.4	33.4	302	10	BB314183	BB314183 BB314183
C 45	50.2	33.2	398	10	BG062030	BG062030 L0955F10-

ALIGNMENTS

RESULT 1	AG180176	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
LOCUS	AG180176	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
DEFINITION	AG180176	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
ACCESSION	AG180176	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
VERSION	AG180176.1	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
KEYWORDS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
SOURCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
ORGANISM	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
TITLE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
JOURNAL	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
REFERENCE	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
AUTHORS	GSS.	Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey	701 bp	DNA	linear	GSS 09-JAN-2002
T						

end was generated during the R&D process and may have higher chance of clone tracking errors.

PRIMERS
Sequencing: T7
LIBRARY

Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI
Location/Qualifiers
1..701
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="RP43-052K19.17"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RPCI-43 Chimpanzee Male BAC Library"

ORIGIN

Query Match 69.4%; Score 104.8; DB 29; Length 701;
Best Local Similarity 98.1%; Pred. No. 6.6e-23;
Matches 106; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 10 CAGGTACCTAAACAGGCCAAATGTCCTTTGGGGTCTCTGTTTCAACAGCATGGTGCA 69
|||||
Db 168 CAGGTACCTAAACAGGCCAAATGTCCTTTGGGGTCTCTGTTTCAACAGCATGGTGCA 227
QY 70 AGCGCGCATCAACCTTCTCGCCATTAAATAAAATGTCATAAACT 117
|||||
Db 228 AGCGTGCATCAACCTTCTCGCCATTAAATAAAATGTCATAAACT 275

RESULT 2

AK050805/c

LOCUS

DEFINITION AK050805 3071 bp mRNA linear HTC 20-SEP-2003
Mus musculus 9 days embryo whole body cDNA, RIKEN full-length
enriched library, clone: D030020M24 product: CORNEAL WOUND HEALING
RELATED PROTEIN homolog [Rattus norvegicus], full insert sequence.

ACCESSION

AK050805.1 GI:26094130

VERSION

HTC; CAP trapper.

KEYWORDS

Mus musculus (house mouse)

SOURCE

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

99279253

MEDLINE

10349636

REFERENCE

2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

20499374

MEDLINE

11042159

REFERENCE

3

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,

Konno, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M.,

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,

Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,

Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer

Genome Res. 10 (11), 1757-1771 (2000)

20530913

MEDLINE

11076861

REFERENCE

4

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

5 The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

6 (bases 1 to 3071)
Adachi, J., Aizawa, K., Akimura, T., Atakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, M., Hiramoto, K., Hirooka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.

Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.

Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers
1..3071
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:D030020M24"
/db_xref="MGI:2418664"
/db_xref="taxon:10090"
/clone="D030020M24"
/tissue type="whole body"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="9 days embryo"

misc_feature
1..3071
/note="CORNEAL WOUND HEALING RELATED PROTEIN homolog
[Rattus norvegicus] (SPTR|Q9J101, evidence: FAST,
96.2%ID, 43.5%length, match=948)"

ORIGIN
Query Match 57.5%; Score 86.8; DB 11; Length 3071;
Best Local Similarity 81.0%; Pred. No. 5.8e-17;
Matches 115; Conservative 0; Mismatches 22; Indels 5; Gaps 1;

QY 10 CAGGTACCTAAACAGGCCAAATGTCCTTTGGGGTCTCTGTTTCAACAGCATGGTGCA 69
|||||
Db 939 CAGGTTCCTAAGCAGGCCAGATGTTGTCCTGAGCTCTCTGTTTCAACACATAGTAGTA 880
QY 70 AGCGCGCATCAACCTTCTCGCCATTAAATAAAATGTCATAAACTCATCTCGCAAGG 129
|||||
Db 879 AGCGTGCATCAACCTTCTCGCTGTCAAAACAATTGTT-----ACTCATCTTGCAGG 825
|||||

QY 130 TGGCAATTCCTCAAGATATG 151
|||||
Db 824 TAGCAACTCTCTCAAGATGTG 803
|||||

```

RESULT 3
CD654321/c
LOCUS
DEFINITION
  CD654321 812 bp mRNA linear EST 18-JUN-2003
  (Long) Homo sapiens cDNA clone IMAGE:30428535 5', mRNA sequence.
ACCESSION
CD654321
KEYWORDS
EST.
SOURCE
  CD654321.1 GI:31893486
  Homo sapiens (human)
ORGANISM
  Homo sapiens
  Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 812)
REFERENCE
  NIH-MGC http://mgc.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Daniela S. Gerhard, Ph.D.
  Office of Cancer Genomics
  National Cancer Institute / NIH
  Bldg. 31 Rm10A07 Bethesda, MD 20892
  Email: cgapbs-x@mail.nih.gov
  Tissue Procurement: Irene Ginis and Mahendra Rao, NIA
  cDNA Library Preparation: Yulan Piao and Minoru Ko
  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Agencourt Bioscience Corporation
  Clone distribution: MGC c lone distribution information
  can be found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: NDAM517 row: 1 column: 16
  High quality sequence stop: 683.
FEATURES
  source
    1..812
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="IMAGE:30428535"
    /tissue_type="Embryonic Stem cells"
    /cell_line="WA01"
    /lab_host="DH10B (T1 phage-resistant)"
    /clone_lib="NIA Human H1 Embryonic Stem Cell cDNA Library
    (Long)"
    /note="vector: pCMV-Sport6; Site 1: NotI; Site 2: SalI;
    This is a long-transcript enriched cDNA library (Genome
    Res. 11: 1553-1558 (2001). [PMID: 11544199]) from WA01
    cell line. Undifferentiated human ES cell line WA01/H1
    was obtained from Wicell Research Institute, Inc.,
    Madison, WI, cultured according to their instructions, on
    MEF feeders. They formed round colonies with defined edges
    and were positive for alkaline phosphatase, SSEA-4, OCT3,
    OCT4, REX1, UTR, SOX2, CX43 and CX45. They are
    negative for GATA2, GATA4, PDX1, NCAM, MSX1, FLT3, SSEA-1,
    TUBB3, NES, GFAP, and EOMES. When confluent (18-10 days
    after plating), the ES cells from 4 X 6cm dishes were
    treated with 1 mg/ml collagenase, type IV
    (Invitrogen/GIBCO) for 5-10 min and gently scraped off
    with 5 ml pipette. RNA was purified with TRIzol Reagent
    from Invitrogen. Protocol ref: Genome Res. 11: 1553-1558
    (2001). [PMID:11544199] Double-stranded cDNAs were
    synthesized with an Oligo(dT) primer [Invitrogen:
    5'-pGACTAGTCTAGATCGAGCGCGCCCTTTTCTTTT-3'] from
    3.4g of total RNA, treated with T4 DNA polymerase, and
    purified by ethanol-precipitation. The cDNAs were ligated
    to lona-linker l1-Sal4, purified by phenol/chloroform
    extraction, and separated from free linkers by
    Centricon-100 column. Then, the cDNAs were amplified by
    long-range high fidelity PCR using Ex Taq polymerase
    (Takara) with a primer Sal4-S for 25 cycles. The products
    were purified by phenol/chloroform extraction and
    Centricon-100 column. The cDNAs were digested with SalI
    and NotI enzymes and cloned into SalI/NotI site of
    pCMV-SPORT6 plasmid vector. The average insert size is
    about 3.6kb."

```

ORIGIN

```

Query Match 49.8%; Score 75.2; DB 14; Length 812;
Best Local Similarity 95.1%; Pred. No. 2.5e-13;
Matches 77; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 37 CTTTGGGGTTCCTGTTTCAACAGCATGGTGTGAAGCGCGCATCAACCTTCTGCTAT 96
Db 730 CTTTGGGGTTCCTGTTTCAACAGCATGGTGTGAAGCGCGCATCAACCTTCTGCTAT 671

QY 97 TAAATAAATAATGTCATAAACT 117
Db 670 TAAATAAATAATGTCATAAACT 650

RESULT 4
CF762228/c
LOCUS
DEFINITION
  CF762228 472 bp mRNA linear EST 17-OCT-2003
  CES000418 Bos taurus skin cDNA library Bos taurus cDNA clone
  CCL000418 5', mRNA sequence.
ACCESSION
CF762228
VERSION
CF762228.1 GI:37711446
KEYWORDS
EST.
SOURCE
  Bos taurus (cow)
ORGANISM
  Bos taurus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
  Bovidae; Bovinae; Bos.
  1 (bases 1 to 472)
REFERENCE
  Wang Y.H., McWilliam S. and Lehnert S.
  Transcription profiling of cattle skin
  Unpublished (2003)
  Contact: Dr Yonghong Wang
  Functional Genomics Lab
  CSIRO Livestock Industries
  Level 5, Queensland Biosciences Precinct, University of Queensland,
  306 Carmody Road St.Lucia QLD Australia
  Tel: 07 3214 2445
  Fax: 07 3214 2685
  Email: Yonghong.Wang@csiro.au
  Plate: 14 row: D column: 09.
FEATURES
  source
    1..472
    /organism="Bos taurus"
    /mol_type="mRNA"
    /strain="Hereford Shorthorn"
    /db_xref="taxon:9913"
    /clone="CCL000418"
    /sex="female"
    /tissue_type="pooled"
    /dev_stage="Adult"
    /lab_host="XLI-BlueMRF'strain"
    /clone_lib="Bos taurus skin cDNA library"
    /notes="Organ: skin; Vector: Uni-ZAPXR; Site 1: EcoRI;
    Site 2: Xho I; Library made from pooled skin of adult
    female Hereford-Shorthorn."
ORIGIN

Query Match 48.9%; Score 73.6; DB 14; Length 472;
Best Local Similarity 91.8%; Pred. No. 6.3e-13;
Matches 78; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 10 CAGTACCTTAACAGGCCAAATGTCCTTGGGGTTCCTGTTTCAACAGCATGGTGA 69
Db 201 CAGTACCTTAACAGGCCAAATGTCCTTGGGGTTCCTGTTTCAACAGCATGGTGA 142

QY 70 AGCGCCGATCAACCTTCTCTGCT 94
Db 141 AGGCTGCATCAACCTTCTCTGCT 117

RESULT 5
BI537770/c

```

LOCUS B1537770 486 bp mRNA linear EST 30-AUG-2001
 DEFINITION 427999 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION B1537770
 VERSION B1537770.1 GI:15378880
 KEYWORDS EST.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus

REFERENCE
 AUTHORS Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G., Perlea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J., and Keele, J.W.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
 1 (bases 1 to 486)
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smithemail.marc.usda.gov
 Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.
 PCR Primers
 FORWARD: AGGAAACAGCTATGACCAT
 BACKWARD: GTTTCCTCAGTCAGCAG
 Plate: 129 row: J column: 14
 Seq primer: ATTAGGTGACACTATAG.
 Location/Qualifiers
 1..486
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /clone_lib="MARC 4BOV"
 /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI; Library made from pooled tissue from day 20 and day 40 embryos."

ORIGIN
 Query Match 48.9%; Score 73.8; DB 12; Length 486;
 Best Local Similarity 91.8%; Pred. No. 6.4e-13;
 Matches 78; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 QY 10 CAGGTACTTAACAGGCCAAATGTTGCTTTGGGGTTCCTGTTTCAACAGCATGGTGGA 69
 Db 278 CAGGTACTTAACAGGCCAAATGTTGCTTTGGGGTTCCTGTTTCAACAGCATAGTATGA 219
 QY 70 AGCGCGCATCAACCTTCTGCGCT 94
 Db 218 AGGCTGCATCAACCTTCTGCGCT 194

RESULT 6
 B1682349/c
 LOCUS B1682349 569 bp mRNA linear EST 17-SEP-2001
 DEFINITION 463408 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION B1682349
 VERSION B1682349.1 GI:15635283
 KEYWORDS EST.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus

REFERENCE
 AUTHORS Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G., Perlea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J., and Keele, J.W.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
 1 (bases 1 to 569)
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smithemail.marc.usda.gov
 Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.
 PCR Primers
 FORWARD: AGGAAACAGCTATGACCAT
 BACKWARD: GTTTCCTCAGTCAGCAG
 Plate: 132 row: B column: 15
 Seq primer: ATTAGGTGACACTATAG.
 Location/Qualifiers
 1..569
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /clone_lib="MARC 1BOV"
 /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI; Library made from pooled tissue from lymph node, ovary, fat, hypothalamus, and pituitary."

ORIGIN
 Query Match 48.9%; Score 73.8; DB 12; Length 569;
 Best Local Similarity 91.8%; Pred. No. 6.6e-13;
 Matches 78; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 QY 10 CAGGTACTTAACAGGCCAAATGTTGCTTTGGGGTTCCTGTTTCAACAGCATGGTGGA 69
 Db 411 CAGGTACTTAACAGGCCAAATGTTGCTTTGGGGTTCCTGTTTCAACAGCATAGTATGA 352
 QY 70 AGCGCGCATCAACCTTCTGCGCT 94
 Db 351 AGGCTGCATCAACCTTCTGCGCT 327

RESULT 7
 B169246
 LOCUS B169246 814 bp mRNA linear EST 30-JAN-2003
 DEFINITION VBB603020215.R1 CSEQFN41 testes Bos taurus cDNA, mRNA sequence.
 ACCESSION B169246
 VERSION B169246.1 GI:28155372
 KEYWORDS EST.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
 1 (bases 1 to 814)
 Adelson, D.L. and Gill, C.A.
 Bovine ESTs (Adelson and Gill)
 Unpublished (2003)
 Contact: David L. Adelson
 Animal Breeding and Genetics
 Texas A&M University
 Animal Science Dept., TAMU-2471, College Station, TX 77843-2471.
 USA
 Tel: 9798452616

REFERENCE
 AUTHORS Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G., Perlea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J., and Keele, J.W.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
 1 (bases 1 to 569)
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smithemail.marc.usda.gov
 Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.
 PCR Primers
 FORWARD: AGGAAACAGCTATGACCAT
 BACKWARD: GTTTCCTCAGTCAGCAG
 Plate: 132 row: B column: 15
 Seq primer: ATTAGGTGACACTATAG.
 Location/Qualifiers
 1..569
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /clone_lib="MARC 1BOV"
 /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI; Library made from pooled tissue from lymph node, ovary, fat, hypothalamus, and pituitary."

ORIGIN
 Query Match 48.9%; Score 73.8; DB 12; Length 569;
 Best Local Similarity 91.8%; Pred. No. 6.6e-13;
 Matches 78; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 QY 10 CAGGTACTTAACAGGCCAAATGTTGCTTTGGGGTTCCTGTTTCAACAGCATGGTGGA 69
 Db 411 CAGGTACTTAACAGGCCAAATGTTGCTTTGGGGTTCCTGTTTCAACAGCATAGTATGA 352
 QY 70 AGCGCGCATCAACCTTCTGCGCT 94
 Db 351 AGGCTGCATCAACCTTCTGCGCT 327

RESULT 7
 B169246
 LOCUS B169246 814 bp mRNA linear EST 30-JAN-2003
 DEFINITION VBB603020215.R1 CSEQFN41 testes Bos taurus cDNA, mRNA sequence.
 ACCESSION B169246
 VERSION B169246.1 GI:28155372
 KEYWORDS EST.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
 1 (bases 1 to 814)
 Adelson, D.L. and Gill, C.A.
 Bovine ESTs (Adelson and Gill)
 Unpublished (2003)
 Contact: David L. Adelson
 Animal Breeding and Genetics
 Texas A&M University
 Animal Science Dept., TAMU-2471, College Station, TX 77843-2471.
 USA
 Tel: 9798452616

Fax: 9798456970
 Email: david.adelson@tamu.edu.

FEATURES
 source
 1. .814
 /location/Qualifiers
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /tissue_type="testes"
 /clone_lib="CSEQFN41 testes"
 /note="Organ: testes; Vector: pBluescript SK+; Site_1: NotI; Site 2: EcoRI; sequence 5' of the insert (5'-NNN...NNNInsert)
 GCAATTGGAGCTCCACGGCGGTGGCGCGCGGCTCGAG. Sequence 3' of the inserts (AAGAAATTCGATACAGCTATGATACCGTGCACCTCGAG. normalized Rd 1 library, sequenced 3' with M13R primer."

ORIGIN
 Query Match 48.9%; Score 73.8; DB 14; Length 814;
 Best Local Similarity 91.8%; Pred. No. 7.1e-13;
 Matches 78; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 10 CAGTACTTAACAGGCCAAATGTTGCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGA 69
 |||||
 Db 134 CAGTACTTAACAGGCCAAATGCTGCTGTGGGGCTCTGTTTCAACAGCATAGTATGA 193
 |||||

QY 70 AGCGCCGATCAACCTTCTCTGCT 94
 |||||
 Db 194 AGGCTGCATCAACCTTCTCTGCT 218
 |||||

RESULT 8
 A2870091
 LOCUS
 DEFINITION 2M0182N22F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0182N22 F, genomic survey sequence.

ACCESSION A2870091
 VERSION A2870091.1 GI:13075085
 KEYWORDS GSS.

SOURCE
 Mus musculus
 Mus musculus (house mouse)

ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 630)

REFERENCE
 AUTHORS
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

TITLE
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL
 COMMENT
 Unpublished (2000)
 Contact: Robert H. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0182 row: N column: 22
 Seq primer: CGTTGTAACGACGCGCAGT
 Class: plasmid ends
 High quality sequence stop: 630.
 Location/Qualifiers
 1. .630
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0182N22"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"

/note="Vector: pWD42ny; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN
 Query Match 48.1%; Score 72.6; DB 28; Length 630;
 Best Local Similarity 85.3%; Pred. No. 1.6e-12;
 Matches 81; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 10 CAGTACTTAACAGGCCAAATGTTGCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGA 69
 |||||
 Db 380 CAGGTTCTTAAACAGGCCAGATGTTGCTCTGAGGGCTCTGTTTCAACACATAGTATGA 439
 |||||

QY 70 AGCGCCGATCAACCTTCTCTGCTATTAAATAA 104
 |||||
 Db 440 AGGCTGCATCAACCTTCTCTGCTGTCACAAACA 474
 |||||

RESULT 9
 BF179099/c
 LOCUS
 DEFINITION 601808136F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:4038945 5', mRNA linear EST 31-OCT-2000
 mRNA sequence.

ACCESSION BF179099
 VERSION BF179099.1 GI:11057241
 KEYWORDS EST.

SOURCE
 Mus musculus (house mouse)

ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 974)

REFERENCE
 AUTHORS
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)

JOURNAL
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-x@mail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LLAM9318 row: g column: 10
 High quality sequence stop: 662.
 Location/Qualifiers
 1. .974
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:4038945"
 /tissue_type="tumor, gross tissue"
 /dev_stage="7 months"
 /lab_host="DH10B"
 /clone_lib="NCI_CGAP Mam5"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;

FEATURES
source

1. .974
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:4038945"
 /tissue_type="tumor, gross tissue"
 /dev_stage="7 months"
 /lab_host="DH10B"
 /clone_lib="NCI_CGAP Mam5"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;

Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

ORIGIN

Query Match 45.3%; Score 68.4; DB 10; Length 974;
Best Local Similarity 76.4%; Pred. No. 4.1e-11;
Matches 84; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 10 CAGGTACCTAAACAGCCCAATGTTGCCCTTGGGTTCTGTTTCAACAGCATGTTGTA 69
|||||
DB 500 CAGGTCTCTAAGCAGCCGACGATGTTCTCTCAGGCTCTCTGTTTCAACACATAGTATGA 441
|||||

QY 70 AGCGCGCATCAACCTTCTCTGCTATTAAATAAAATGTCATAACTCA 119
|||||
DB 440 AGCGTGCATCAACCTTCTCTGCCCACTTTCAGGTAGTAGCACTCTCA 391
|||||

RESULT 10
LOCUS AI020180/c
DEFINITION ubi5b06.r1 Stratagene mouse macrophage (#937306) Mus musculus cDNA
clone IMAGE:1367027 5', mRNA sequence.

ACCESSION AI020180
VERSION AI020180.1 GI:3234516
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:900247

Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 397.

FEATURES

source

1..409
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:1367027"
/tissue_type="macrophage"
/dev_stage="WEHI-3 cell line"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="Stratagene mouse macrophage (#937306)"
/note="Organ: blood; Vector: pBluescript SK-; Site 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. WEHI-3 cell line. Average insert size: 1.5 kb;
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGG
3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."

ORIGIN

Query Match 44.6%; Score 67.4; DB 9; Length 409;
Best Local Similarity 87.1%; Pred. No. 7.2e-11;
Matches 74; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 10 CAGGTACCTAAACAGCCCAATGTTGCCCTTGGGTTCTCTTCAACAGCATGTTGTA 69

Db 380 CAGGTCTCTAAGCAGCCGACGATGTTGTTCTCAGGCTCTCTTCAACATAGTATGA 321
QY 70 AGCGCGCATCAACCTTCTCTGCGCT 94
DB 320 AGCGTGCATCAACCTTCTCTGCGCT 296

RESULT 11
LOCUS BB727790/c
DEFINITION BB727790.1 GI:16111065
ACCESSION BB727790
VERSION BB727790.1
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE

AUTHORS 1 (bases 1 to 426)

Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.

JOURNAL Unpublished (2001)

COMMENT

Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.

FEATURES

source

1..426
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="E860029E01"
/cell_type="8 cells"
/dev_stage="8 cells embryo"
/clone_lib="RIKEN full-length enriched, 8 cells embryo"

ORIGIN

```

Query Match      44.6%; Score 67.4; DB 10; Length 426;
Best Local Similarity 87.1%; Pred. No. 7.3e-11;
Matches 74; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 10 CAGGTACCTAAACAGGCGCAATGTTGCTTTGGGGTTCCTGTTTCAACACAGCATGGTGTGA 69
    |||||
Db 218 CAGGTTCTTAAGCAGGCGCAGATGTTGCTCTGAGGCTCTCTGTTTCAACACATAGTAIGA 159
    |||||

QY 70 AGCGCGCATCAACCTTCTCTGCT 94
    |||||
Db 158 AGCGCTGCATCAACCTTCTCTGCT 134

RESULT 12
BB827250/c
LOCUS BB827250
DEFINITION BB827250 RIKEN full-length enriched, mammary gland RCB-0526
JYG-MC(A) cDNA Mus musculus cDNA clone G830046013 3', mRNA
sequence.
EST. BB827250
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 435)
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hirakawa,T., Hirozane,T., Imotani,K.,
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akaira,S.,
Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp/
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
FEATURES
Location/Qualifiers
1. 435
/organism="Mus musculus"
/mol_type="mRNA"

```

```

/db_xref="taxon:10090"
/clone="G830046013"
/tissue_type="mammary gland"
/cell_line="RCB-0526 Jyg-MC(A)"
/clone_lib="RIKEN full-length enriched, mammary gland
RCB-0526 Jyg-MC(A) cDNA"

ORIGIN

Query Match      44.6%; Score 67.4; DB 10; Length 435;
Best Local Similarity 87.1%; Pred. No. 7.3e-11;
Matches 74; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 10 CAGGTACCTAAACAGGCGCAATGTTGCTTTGGGGTTCCTGTTTCAACACAGCATGGTGTGA 69
    |||||
Db 227 CAGGTTCTTAAGCAGGCGCAGATGTTGCTCTGAGGCTCTCTGTTTCAACACATAGTAIGA 168
    |||||

QY 70 AGCGCGCATCAACCTTCTCTGCT 94
    |||||
Db 167 AGCGCTGCATCAACCTTCTCTGCT 143

RESULT 13
BB832780/c
LOCUS BB832780
DEFINITION BB832780 RIKEN full-length enriched, mammary gland RCB-0527
JYG-MC(B) cDNA Mus musculus cDNA clone G930023L04 3', mRNA
sequence.
EST. BB832780
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 446)
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hirakawa,T., Hirozane,T., Imotani,K.,
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akaira,S.,
Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp/
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
FEATURES
Location/Qualifiers
1. 435
/organism="Mus musculus"
/mol_type="mRNA"

```

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

source
Location/Qualifiers
1..446
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="G930023L04"
/tissue type="mammary gland"
/cell_line="RCB-0527 Jyg-MC(B)"
/clone_lib="RIKEN full-length enriched, mammary gland
RCB-0527 Jyg-MC(B) cDNA"

ORIGIN

Query Match 44.6%; Score 67.4; DB 10; Length 446;
Best Local Similarity 87.1%; Pred. No. 7.3e-11;
Matches 74; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 10 CAGGTACCTAAACAGGCCAAATGTTGCCCTTTGGGGTTCCTGTTTCAACAGCATGTGTGA 69
|||||
Db 232 CAGGTCTCTAAGCAGGCCAGATGTTGCTCTGAGGCTCTCTGTTTCAACACATAGTATGA 173
|||||
QY 70 AGCGCGCATCAACCTTCTCTGCCT 94
|||||
Db 172 AGCGTGCATCAACCTTCTCTGCCT 148
|||||

RESULT 14
BE308753/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BE308753 478 bp mRNA linear EST 26-OCT-2000
601091825F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3486246 5',
mRNA sequence.

BE308753
BE308753.1 GI:9165894

EST.
Mus musculus (house mouse)

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at:

<http://image.llnl.gov>

Plate: LLM8322 row: b column: 07

High quality sequence stop: 478.

Location/Qualifiers

FEATURES

source
1..478
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3486246"
/tissue type="tumor, gross tissue"
/dev stage="7 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP Mam5"

/note="Organ: Mammary; Vector: pCMV-SPORT6; Site1: SalI;
Site2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

ORIGIN

Query Match 44.6%; Score 67.4; DB 14; Length 512;
Best Local Similarity 87.1%; Pred. No. 7.6e-11;

Query Match 44.6%; Score 67.4; DB 10; Length 478;
Best Local Similarity 87.1%; Pred. No. 7.4e-11;
Matches 74; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 10 CAGGTACCTAAACAGGCCAAATGTTGCCCTTTGGGGTTCCTGTTTCAACAGCATGTGTGA 69
|||||
Db 315 CAGGTCTCTAAGCAGGCCAGATGTTGCTCTGAGGCTCTCTGTTTCAACACATAGTATGA 256
|||||
QY 70 AGCGCGCATCAACCTTCTCTGCCT 94
|||||
Db 255 AGCGTGCATCAACCTTCTCTGCCT 231
|||||

RESULT 15

W54490/c

LOCUS

DEFINITION
W54490 512 bp mRNA linear EST 03-JUN-1996
md09d10.r1 Soares mouse embryo NME13.5 14.5 Mus musculus cDNA
clone IMAGE:367891 5', mRNA sequence.

ACCESSION
W54490.1 GI:1355550

VERSION

KEYWORDS

SOURCE
Mus musculus (house mouse)

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 512)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mousest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:229323

Seq primer: RTPrimer

High quality sequence stop: 344.

Location/Qualifiers

1..512

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:367891"

/sex="unknown"

/tissue type="embryo"

/dev stage="13.5-14.5dpc total fetus"

/lab_host="DH10B"

/clone_lib="Soares mouse embryo NME13.5 14.5"

/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M. Patricia Bonaldo."

ORIGIN

Query Match 44.6%; Score 67.4; DB 14; Length 512;
Best Local Similarity 87.1%; Pred. No. 7.6e-11;

Matches	74;	Conservative	0;	Mismatches	11;	Indels	0;	Gaps	0;
QY	10	CAGGTACCTAAACAGGCCAAATGTGCTTTGGGTTCCCTGTTTCAACAGCATGGTGTGA	69						
Db	404	CAGGTTCTTAAGCAGGCCAGATGTTGTCTCTGAGGCTCCTGTTTCAACACATAGTATGA	345						
QY	70	AGGCGCGCATCAACCTTCTCTGCCT	94						
Db	344	AGGCTGCATCAACCTTCTCTGCCT	320						

Search completed: August 11, 2004, 04:18:35
 Job time : 415.913 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 10, 2004, 17:10:17 ; Search time 60.5488 Seconds
(without alignments)
10594.393 Million cell updates/sec

Title: US-10-001-857-41

Perfect score: 151
Sequence: 1 ccgccggcaggtagctaccta.....gcaattctcagaatatg 151

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_29Jan04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002s:*

7: Geneseqn2003as:*

8: Geneseqn2003bs:*

9: Geneseqn2003cs:*

10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	151	100.0	151	6	ABQ75302 Human lun
2	140.4	93.0	3096	6	ABQ75303 Human lun
3	84.8	56.2	2488	6	AAL49929 Human mol
4	83.8	55.5	303	5	ABV06162 Human pro
5	83.8	55.5	439	5	ABV45182 Human pro
6	83.8	55.5	439	5	ABV36125 Human pro
7	83.8	55.5	815	6	ABQ75372 Human lun
8	83.2	55.1	2668	5	ABV29823 Human pro
9	83.2	55.1	2668	5	ABV23940 Human pro
10	83.2	55.1	3029	9	ADC30041 Human nov
11	82.8	54.8	356	5	ABV15331 Human pro
12	63.2	41.9	2543	9	ADB53632 Primary r
13	33.4	22.1	3215	6	ABL55204 HBV subty
14	32.2	21.3	129021	3	AAF22296 BAC conta
15	31.8	21.1	560	1	ABT23617 Stabillisi
16	31.8	21.1	3214	1	AAN60714 Sequence
17	31.8	21.1	160755	4	AAB88704 Human DNA
18	31	20.5	1445	2	AAH8684 Fulminant
19	31	20.5	1500	2	AAV82693 Fulminant
20	31	20.5	1644	7	ACA35581 Prokaryot
21	30.8	20.4	362	5	ABV32031 Human pro
22	30.8	20.4	487	5	ABV40970 Human pro
23	30.6	20.3	322101	9	AAD58431 Human PAO

C	24	30.2	20.0	462	2	AAT05544	Rat05544 Human hep
C	25	30.2	20.0	560	7	ABT23618	Abt23618 Stabillisi
C	26	30.2	20.0	3214	2	AAQ05377	Aaq05377 Fragment
C	27	30.2	20.0	3835	2	AAQ05378	Aaq05378 Sequence
C	28	30.2	20.0	4421	2	AAT58319	Rat58319 DNA sequ
C	29	30	19.9	13273	4	AAS36849	Aas36849 Human car
C	30	30	19.9	13273	9	ADE47543	Ade47543 Human car
C	31	30	19.9	35057	7	AAL51501	Rat51501 Human pan
C	32	30	19.9	300000	9	ADE86352	Ade86352 Human PTP
C	33	29.6	19.6	576	7	ABZ39816	Abz39816 N. gonorr
C	34	29.6	19.6	576	7	ACA41490	Ac41490 Prokaryot
C	35	29.6	19.6	1035	3	AZ54048	Aaz54048 Neisseria
C	36	29.4	19.5	197	6	AAS16091	Aas16091 Hepatitis
C	37	29.4	19.5	250	6	ABK29867	Abk29867 Wild type
C	38	29.4	19.5	1445	2	AAV82692	AAV82692 Fulminant
C	39	29.4	19.5	1445	2	AAV82685	AAV82685 Fulminant
C	40	29.4	19.5	1445	2	AAV82690	AAV82690 Fulminant
C	41	29.4	19.5	1500	2	AAV82694	AAV82694 Fulminant
C	42	29.4	19.5	1500	2	AAV82697	AAV82697 Fulminant
C	43	29.4	19.5	1500	2	AAV82686	AAV82686 Fulminant
C	44	29.4	19.5	2342	1	AAN93072	Aan93072 Sequence
C	45	29.4	19.5	3182	6	AAD31765	Aad31765 Hepatitis

ALIGNMENTS

RESULT 1

ABQ75302

ID ABQ75302 standard; cdna; 151 BP.

AC ABQ75302;

XX

XX 05-NOV-2002 (first entry)

XX

XX Human lung specific nucleic acid sequence SEQ ID NO:41.

XX

XX Human; lung; lung specific nucleic acid; LSNA; lung specific protein;
XX LSP; cytostatic; gene therapy; vaccine; metastasis; lung cancer;
XX squamous cell carcinoma; gene; chromosome 9; ss.

XX

XX Homo sapiens.

XX

XX WO200264788-A2.

XX

XX 22-AUG-2002.

XX

XX 20-NOV-2001; 2001WO-US045080.

XX

XX 20-NOV-2000; 2000US-0252054P.

XX

XX (DIAD-) DIADEXUS INC.

XX

XX Macina RA, Recipon H, Chen S, Sun Y, Liu C;

XX

XX WPI; 2002-657601/70.

XX

XX New lung specific nucleic acid useful in gene therapy or as vaccines for
XX treating lung cancer (e.g. squamous cell carcinoma) or non-carcinoma lung
XX diseases, as well as for diagnosing, monitoring or staging these
XX diseases.

XX

XX Claim 1; Page 168; 282pp; English.

XX

XX The present invention describes an isolated lung specific nucleic acid
XX (LSNA) comprising a sequence that: (a) encodes any of the 93 amino acid
XX sequences comprising 17 - 733 amino acids, given in ABP52873 to ABP52965;
XX (b) comprises any of 115 sequences comprising 148 - 3193 base pairs (bp),
XX given in ABQ75262 to ABQ75376; (c) selectively hybridizes to (a) or (b);
XX or (d) has 60 % sequence identity to (a) or (b). LSNA and lung specific
XX protein (LSP) sequences have cytostatic activity and can be used in gene
XX therapy and vaccines. LSNA and LSPs are useful for diagnosing and
XX monitoring the presence and metastases of lung cancer in a patient. An

CC antibody that specifically binds to an LSP can be used for determining
 CC the presence of an LSP in a sample, as well as for treating a patient
 CC with lung cancer, particularly by inducing an immune response against the
 CC lung cancer cell expressing the LSNAs or LSPs. In particular, these LSNAs
 CC and LSPs are useful for identifying, diagnosing, monitoring, staging,
 CC imaging and treating lung cancer (e.g. squamous cell carcinoma) and non-
 CC cancerous disease states in lung

XX Sequence 151 BP; 42 A; 39 C; 32 G; 38 T; 0 U; 0 Other;

Query Match 100.0%; Score 151; DB 6; Length 151;
 Best Local Similarity 100.0%; Pred. No. 3.6e-43;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGCCCGGCGAGTACCTAAACAGGCCAAATGTCCTTTGGGGTTCCTGTTTCAACAGC 60
 DB 1 CCGCCCGGCGAGTACCTAAACAGGCCAAATGTCCTTTGGGGTTCCTGTTTCAACAGC 60
 QY 61 ATGGTGTGAAGCGCGCATCAACCTTCTGCTATTAAATATAAATGTCATAAATCAT 120
 DB 61 ATGGTGTGAAGCGCGCATCAACCTTCTGCTATTAAATATAAATGTCATAAATCAT 120
 QY 121 CTGCAAGTGGCAAAATTCCTCAAGATATG 151
 DB 121 CTGCAAGTGGCAAAATTCCTCAAGATATG 151

RESULT 2

ABQ75303/c
 ID ABQ75303 standard; cDNA; 3096 BP.

XX AC ABQ75303;

XX DT 05-NOV-2002 (first entry)

XX Human lung specific nucleic acid sequence SEQ ID NO:42.

XX Human; lung; lung specific nucleic acid; LSNA; lung specific protein;
 KW LSP; cytostatic; gene therapy; vaccine; metastasis; lung cancer;
 KW squamous cell carcinoma; gene; chromosome 9; ss.

XX OS Homo sapiens.

XX PN WO200264788-A2.

XX PD 22-AUG-2002.

XX PF 20-NOV-2001; 2001WO-US045080.

XX PR 20-NOV-2000; 2000US-0252054P.

XX PA (DIAD-) DIADEXUS INC.

XX PI Macina RA, Recipon H, Chen S, Sun Y, Liu C;

XX DR WPI; 2002-657601/70.

XX New lung specific nucleic acid useful in gene therapy or as vaccines for
 PT treating lung cancer (e.g. squamous cell carcinoma) or non-cancerous lung
 PT diseases, as well as for diagnosing, monitoring or staging these
 PT diseases.

XX PS Claim 1; Page 168-169; 282pp; English.

XX The present invention describes an isolated lung specific nucleic acid
 CC (LSNA) comprising a sequence that: (a) encodes any of the 93 amino acid
 CC sequences comprising 17 - 733 amino acids, given in ABP52873 to ABP52965;
 CC (b) comprises any of 115 sequences comprising 148 - 3193 base pairs (bp),
 CC given in ABQ75262 to ABQ75376; (c) selectively hybridises to (a) or (b);
 CC or (d) has 60 % sequence identity to (a) or (b). LSNA and lung specific
 CC protein (LSP) sequences have cytostatic activity and can be used in gene
 CC therapy and vaccines. LSNAs and LSPs are useful for diagnosing and
 CC monitoring the presence and metastases of lung cancer in a patient. An

CC antibody that specifically binds to an LSP can be used for determining
 CC the presence of an LSP in a sample, as well as for treating a patient
 CC with lung cancer, particularly by inducing an immune response against the
 CC lung cancer cell expressing the LSNAs or LSPs. In particular, these LSNAs
 CC and LSPs are useful for identifying, diagnosing, monitoring, staging,
 CC imaging and treating lung cancer (e.g. squamous cell carcinoma) and non-
 CC cancerous disease states in lung

XX Sequence 3096 BP; 914 A; 606 C; 749 G; 827 T; 0 U; 0 Other;

Query Match 93.0%; Score 140.4; DB 6; Length 3096;
 Best Local Similarity 99.3%; Pred. No. 6.9e-39;
 Matches 141; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 10 CAGGTACTTAAACAGGCCAAATGTTGCTTTGGGGTTCCTGTTTCAACAGCATGTGTGA 69
 DB 2052 CAGGTACTTAAACAGGCCAAATGTTGCTTTGGGGTTCCTGTTTCAACAGCATGTGTGA 1993
 QY 70 AGCGCGCATCAACCTTCTGCTATTAAATATAAATGTCATAAATCATCTCTCAAGG 129
 DB 1992 AGCGTGCATCAACCTTCTGCTATTAAATATAAATGTCATAAATCATCTCTCAAGG 1993
 QY 130 TGGCAATTCCTCAAGATATG 151
 DB 1932 TGGCAATTCCTCAAGATATG 1911

RESULT 3

AA49929/c

ID AA49929 standard; cDNA; 2488 BP.

XX AC AA49929;

XX DT 10-DEC-2002 (first entry)

XX Human molecule for disease detection and treatment coding sequence #3.

XX Human; molecule for disease detection and treatment; MDDT; gene therapy;
 KW cytostatic; antiarteriosclerotic; hepatotropic; anti-HIV; antiallergic;
 KW antiinflammatory; antiaesthetic; cerebroprotective; nootropic;
 KW neuroprotective; antiparkinsonian; cardiac; antianginal; gene; ss.

XX OS Homo sapiens.

XX PN WO200270709-A2.

XX PD 12-SEP-2002.

XX PF 08-FEB-2002; 2002WO-US003709.

XX PR 09-FEB-2001; 2001US-0268117P.

XX PR 15-FEB-2001; 2001US-0269618P.

XX PR 23-FEB-2001; 2001US-0271118P.

XX PR 07-MAR-2001; 2001US-0274486P.

XX PR 09-MAR-2001; 2001US-0274436P.

XX PR 28-NOV-2001; 2001US-0334229P.

XX PR 01-FEB-2002; 2002US-0353284P.

XX PA (INCY-) INCYTE GENOMICS INC.

XX WPI; 2002-713453/77.

XX P-PSDB; AAO19400.

XX New human molecules for disease detection and treatment (MDDT), useful
 PT for diagnosing, treating and preventing diseases or conditions associated
 PT with the aberrant MDDT expression, e.g. cancer, AIDS, asthma, diabetes,
 PT hepatitis.

Lal PG, Baughn MR, Yao MG, Wallia NK, Elliot VS, Xu Y;
 Honchell CD, Yue H, Ding L, Gietzen KJ, Ison CH, Lu DM;
 Hafalia AJA, Ghadi AR, Thangavelu K, Sanjanwala WM, Tang YT;
 Rankumar J, Griffin JA, Swarnaker A, Azimzai Y, Sapperstein SK;
 Burford N, Lee EA, Lu Y, Tran UK, Marquis JP;

XX Claim 11; Page 159-160; 177pp; English.

XX The present invention relates to human proteins and coding sequences of

CC molecules for disease detection and treatment MDDT. The sequences can be

CC used in the treatment of diseases associated with the decreased

CC expression or overexpression of MDDT, such as cell proliferative (cancer,

CC atherosclerosis, hepatitis), autoimmune/inflammatory (e.g. AIDS,

CC allergies, Addison's disease, asthma), developmental (dwarfism, renal

CC tubular acidosis), neurological (e.g. stroke, Parkinson's disease,

CC epilepsy) and cardiovascular (congestive heart failure, myocardial

CC infarction, angina pectoris) disorders. The present sequence is a coding

CC sequence of the invention

XX Sequence 2488 BP; 768 A; 461 C; 567 G; 692 T; 0 U; 0 Other;

Query Match 56.2%; Score 84.8; DB 6; Length 2488;

Best Local Similarity 88.5%; Pred. No. 2.8e-19;

Matches 92; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 10 CAGGTACCTAAACAGGCCAAATGTCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGA 69

Db 1554 CAGGTACCTAAACAGGCCAAATGTCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGA 1495

QY 70 AGCGCGCATCAACCTTCTGCTGCTATTAAATAAATATGTCATA 113

Db 1494 AGCGGCTCATCAACCTTCTGCTGCTCATCTGCTCAAGGTGGCAAA 1451

RESULT 4

ABV06162

ID ABV06162 standard; cDNA; 303 BP.

XX AC ABV06162;

XX DT 13-SEP-2002 (first entry)

XX DE Human prostate expression marker cDNA 6153.

XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

XX pharmacogenomic marker; gene; ss.

XX Homo sapiens.

XX WO200160860-A2.

XX 23-AUG-2001.

XX 20-FEB-2001; 2001WO-US005171.

FR 17-FEB-2000; 2000US-0183319P.

FR 16-MAR-2000; 2000US-0189862P.

PR 25-MAY-2000; 2000US-0207454P.

PR 09-JUN-2000; 2000US-0211314P.

PR 18-JUL-2000; 2000US-0219007P.

PR 13-DEC-2000; 2000US-0255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Endege WO, Monahan JE;

XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of

PT prostate cells and correlating with presence of prostate cancer, useful

PT for detecting presence of prostate cancer, stage of prostate cancer.

XX Claim 1; Page 1018; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising

CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the

CC specification or its complement. (I) is useful for: (a) assessing whether

CC a patient is afflicted with prostate cancer; (b) monitoring the

XX progression of prostate cancer in a patient; (c) assessing the efficacy

CC of a test compound to inhibit prostate cancer in a patient; (d) assessing

CC the efficacy of a therapy for inhibiting prostate cancer in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)

CC progression of prostate cancer in a patient; (c) assessing the efficacy

CC of a test compound to inhibit prostate cancer in a patient; (d) assessing

CC the efficacy of a therapy for inhibiting prostate cancer in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)

CC determining whether prostate cancer has metastasized in a patient; (h)

CC assessing the aggressiveness or indolence of prostate cancer in a patient

CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker

XX Sequence 303 BP; 81 A; 72 C; 77 G; 73 T; 0 U; 0 Other;

Query Match 55.5%; Score 83.8; DB 5; Length 303;

Best Local Similarity 88.3%; Pred. No. 2.7e-19;

Matches 91; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 11 AGGTACCTAAACAGGCCAAATGTCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGA 70

Db 22 AGGTACCTAAACAGGCCAAATGTCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGA 81

QY 71 GCGCCGCATCAACCTTCTGCTGCTATTAAATAAATATGTCATA 113

Db 82 GCGCTGCATCAACCTTCTGCTGCTCATCTGCTCAAGGTGGCAAA 124

RESULT 5

ABV45182

ID ABV45182 standard; cDNA; 439 BP.

XX AC ABV45182;

XX DT 16-SEP-2002 (first entry)

XX DE Human prostate expression marker cDNA 45173.

XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

XX pharmacogenomic marker; gene; ss.

XX Homo sapiens.

XX WO200160860-A2.

XX 23-AUG-2001.

XX 20-FEB-2001; 2001WO-US005171.

FR 17-FEB-2000; 2000US-0183319P.

PR 16-MAR-2000; 2000US-0189862P.

PR 25-MAY-2000; 2000US-0207454P.

PR 09-JUN-2000; 2000US-0211314P.

PR 18-JUL-2000; 2000US-0219007P.

PR 13-DEC-2000; 2000US-0255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Endege WO, Monahan JE;

XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of

PT prostate cells and correlating with presence of prostate cancer, useful

PT for detecting presence of prostate cancer, stage of prostate cancer.

XX Claim 1; Page 8947; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising

CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the

CC specification or its complement. (I) is useful for: (a) assessing whether

CC a patient is afflicted with prostate cancer; (b) monitoring the

CC progression of prostate cancer in a patient; (c) assessing the efficacy

CC of a test compound to inhibit prostate cancer in a patient; (d) assessing

CC the efficacy of a therapy for inhibiting prostate cancer in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)

CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 439 BP; 113 A; 104 C; 110 G; 112 T; 0 U; 0 Other;

Query Match 55.5%; Score 83.8; DB 5; Length 439;
Best Local Similarity 88.3%; Pred. No. 3.2e-19;
Matches 91; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 11 AGGTACCTAAACAGCCAAATGTTGCCCTTGGGGTTCCTGTTTCAACAGCATGGTGTGAA 70
Db 100 AGGTACCTAAACAGCCAAATGTTGCCCTTGGGGTTCCTGTTTCAACAGCATGGTGTGAA 159

QY 71 GCGCGCATCAACCTTCTGCTATTAATAAATATGTCATA 113
Db 160 GCGCTGCATCAACCTTCTGCTATTCCTCAAGGTGGCAAA 202

RESULT 6
ABV36125
ID ABV36125 standard; cDNA; 439 BP.
XX
AC ABV36125;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 36116.

XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.

PN WO200160860-A2.
XX
PD 23-AUG-2001.

PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.

PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Endege WO, Monahan JE;

XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX

PS Claim 1; Page 7483; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker

SQ Sequence 439 BP; 113 A; 104 C; 110 G; 112 T; 0 U; 0 Other;

Query Match 55.5%; Score 83.8; DB 5; Length 439;
Best Local Similarity 88.3%; Pred. No. 3.2e-19;
Matches 91; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 11 AGGTACCTAAACAGCCAAATGTTGCCCTTGGGGTTCCTGTTTCAACAGCATGGTGTGAA 70
Db 100 AGGTACCTAAACAGCCAAATGTTGCCCTTGGGGTTCCTGTTTCAACAGCATGGTGTGAA 159

QY 71 GCGCGCATCAACCTTCTGCTATTAATAAATATGTCATA 113
Db 160 GCGCTGCATCAACCTTCTGCTATTCCTCAAGGTGGCAAA 202

RESULT 7
ABQ75372
ID ABQ75372 standard; cDNA; 815 BP.
XX

AC ABQ75372;

XX 05-NOV-2002 (first entry)

XX Human lung specific nucleic acid sequence SEQ ID NO:111.

XX Human; lung; lung specific nucleic acid; LSNA; lung specific protein;
KW LSP; cytostatic; gene therapy; vaccine; metastasis; lung cancer;
KW squamous cell carcinoma; gene; chromosome 9; ss.

XX Homo sapiens.

PN WO200264788-A2.

XX 22-AUG-2002.

PF 20-NOV-2001; 2001WO-US045080.

PR 20-NOV-2000; 2000US-0252054P.

XX (DIAD-) DIADEXUS INC.

XX Macina RA, Recipon H, Chen S, Sun Y, Liu C;

XX WPI; 2002-657601/70.

XX New lung specific nucleic acid useful in gene therapy or as vaccines for
PT treating lung cancer (e.g. squamous cell carcinoma) or non-cancerous lung
PT diseases, as well as for diagnosing, monitoring or staging these
PT diseases.

PS Claim 1; Page 223; 282pp; English.

XX The present invention describes an isolated lung specific nucleic acid
CC (LSNA) comprising a sequence that: (a) encodes any of the 93 amino acid
CC sequences comprising 17 - 733 amino acids, given in ABP52873 to ABP52965;
CC (b) comprises any of 115 sequences comprising 148 - 3193 base pairs (bp),
CC given in ABQ75262 to ABQ75376; (c) selectively hybridizes to (a) or (b);
CC or (d) has 60 % sequence identity to (a) or (b). LSNA and lung specific
CC protein (LSP) sequences have cytostatic activity and can be used in gene
CC therapy and vaccines. LSNA and LSPs are useful for diagnosing and
CC monitoring the presence and metastases of lung cancer in a patient. An
CC antibody that specifically binds to an LSP can be used for determining
CC the presence of an LSP in a sample, as well as for treating a patient
CC with lung cancer, particularly by inducing an immune response against the
CC lung cancer cell, expressing the LSNA or LSPs. In particular, these LSNA
CC and LSPs are useful for identifying, diagnosing, monitoring, staging,
CC imaging and treating lung cancer (e.g. squamous cell carcinoma) and non-
CC cancerous disease states in lung

SQ Sequence 815 BP; 310 A; 116 C; 207 G; 82 T; 0 U; 100 Other;

Query Match 55.5%; Score 83.8; DB 6; Length 815;
Best Local Similarity 88.3%; Pred. No. 4.1e-19;
Matches 91; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

[illegible]

RESULT 8	
ABV29823	
ID	ABV29823 standard; cDNA; 2668 BP.
XX	
XX	
AC	ABV29823;
XX	
XX	
DT	16-SEP-2002 (first entry)
XX	
DE	Human prostate expression marker cDNA 29814.
XX	
XX	
KW	Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW	pharmacogenomic marker; gene; ss.
KW	

Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.

Claim 1; Page 6420; 11750pp; English.

The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker

Sequence 2668 BP: 742 A; 617 C; 483 G; 820 T; 0 U; 6 Other

	70	ASGGCGCATCAACTCTTCTGTCCTATTAAAAATAAATGTGCATA	115
Qy			
Dd	1127	AGSGTGTCATCAACCTTCTGTCCATCCTGCAAGGTGGCAA	1170
 RESULT 9 ABV23940 ID ABV23940 standard; cDNA; 2668 BP.			

RESULT 9	
ABV23940	
ID	ABV23940 standard; cDNA; 2668 BP.
XX	
XX	
AC	ABV23940;
XX	
XX	
DT	16-SEP-2002 (first entry)
XX	
XX	
DE	Human prostate expression marker CDNA 23931.
XX	
KW	Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW	pharmacogenomic marker; gene; ss.
XX	
OS	Homo sapiens.

xx Novel isolated nucleic acid molecule associated with cancerous state of
pt prostate cells and correlating with presence of prostate cancer, useful
pt for detecting presence of prostate cancer, stage of prostate cancer.
pt

The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient; (i) is also useful as a pharmacodynamic or pharmacogenomic marker ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

Qy
70 AGCGCGCATCAACCTTCTCTGCCTATTAATAAATAAAATGTCATA 113

pB
1127 AGCGTGTCATCAACCTTCTCTGCCTATCTGTCAAGTGGCAA 1170

CC claimed human cDNA sequence of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 3029 BP; 895 A; 593 C; 739 G; 802 T; 0 U; 0 Other;

Query Match 55.1%; Score 83.2; DB 9; Length 3029;
Best Local Similarity 87.5%; Pred. No. 1.1e-18;
Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 10 CAGGTACCTTAACAGCGCAAATGTTGCCCTTGGGGTTCCTGTTTCAACAGCATGGTGGA 69
| | | | |
Db 1983 CAGGTACCTTAACAGCGCAAATGTTGCCCTTGGGGTTCCTGTTTCAACAGCATGGTGGA 69
| | | | |
QY 70 AGCGCGGCATCAACCTTCTCTGCTATTAAATAAATGTCATA 113
| | | | |
Db 1923 AGCGCTGCATCAACCTTCTCTGCTCATCTCGCAGGTGGCAA 1880
| | | | |

RESULT 11
ABV15331:
ID ABV15331 standard; cDNA; 356 BP.

XX AC ABV15331;
XX
DT 13-SEP-2002 (first entry)
XX
XX Human prostate expression marker cDNA 15322.
XX
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.
XX Homo sapiens.
XX OS
XX WO200160860-A2.
XX PN
XX PD 23-AUG-2001.
XX PP
XX 20-FEB-2001; 2001WO-US005171.
XX
XX 17-FEB-2000; 2000US-0183319P.
XX PR 16-MAR-2000; 2000US-0189862P.
XX PR 25-MAY-2000; 2000US-0207454P.
XX PR 09-JUN-2000; 2000US-0211314P.
XX PR 18-JUL-2000; 2000US-0219007P.
XX PR 13-DEC-2000; 2000US-0255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PA
XX Schlegel R, Endege WO, Monahan JB;
XX PI WPI; 2001-662795/76.
XX DR
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful
XX for detecting presence of prostate cancer, stage of prostate cancer.
XX
XX Claim 1; Page 2572; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX specification or its complement. (I) is useful for: (a) assessing whether
XX a patient is afflicted with prostate cancer; (b) monitoring the
XX progression of prostate cancer in a patient; (c) assessing the efficacy
XX of a test compound to inhibit prostate cancer in a patient; (d) assessing
XX the efficacy of a therapy for inhibiting prostate cancer in a patient;
XX (e) selecting a composition for inhibiting prostate cancer in a patient;
XX (f) assessing the prostate cell carcinogenic potential of a compound; (g)
XX determining whether prostate cancer has metastasized in a patient; (h)
XX assessing the aggressiveness or indolence of prostate cancer in a patient
XX ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker


```
SQ Sequence 356 BP; 97 A; 80 C; 85 G; 94 T; 0 U; 0 Other;
Query Match 54.8%; Score 82.8; DB 5; Length 356;
Best Local Similarity 88.2%; Pred. No. 6.6e-19;
Matches 90; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 12 GGTACCTAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGCATGTTGTAAG 71
DB 1 GGTACCTAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGCATGTTGTAAG 60
QY 72 CGCGCATCAACCTTCTCTGCTATTATAAATAAAATGTCATA 113
DB 61 CGCTGATCAACCTTCTCTGCTATCACTCGCTCAAGGTGGCAA 102
RESULT 12
ADB53632/c
ID ADB53632 standard; DNA; 2543 BP.
XX
AC ADB53632;
XX
DT 04-DEC-2003 (first entry)
DE Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:4174.
XX
KW toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;
KW toxicity marker; toxicity progression; drug screening;
KW primary rat hepatocyte toxicity modelling; gene; ds.
XX
OS Rattus norvegicus.
XX
FN WO2003065993-A2.
XX
PD 14-AUG-2003.
XX
PX 04-FEB-2003; 2003WO-US003482.
XX
PR 04-FEB-2002; 2002US-0353171P.
PR 13-MAR-2002; 2002US-0363534P.
PR 08-APR-2002; 2002US-0370248P.
PR 10-APR-2002; 2002US-0371134P.
PR 10-APR-2002; 2002US-0371135P.
PR 10-APR-2002; 2002US-0371150P.
PR 11-APR-2002; 2002US-0371413P.
PR 19-APR-2002; 2002US-0373601P.
PR 19-APR-2002; 2002US-0373602P.
PR 22-APR-2002; 2002US-0374139P.
PR 08-MAY-2002; 2002US-0378370P.
PR 09-MAY-2002; 2002US-0378652P.
PR 09-MAY-2002; 2002US-0378653P.
PR 09-MAY-2002; 2002US-0378655P.
PR 09-JUL-2002; 2002US-0394230P.
PR 09-JUL-2002; 2002US-0394253P.
PR 04-SEP-2002; 2002US-0407688P.
PR 28-JAN-2003; 2003US-0442900P.
XX
PA (GENE-) GENE LOGIC INC.
XX
XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;
PI Elashoff M;
XX
DR WPI; 2003-731472/69.
XX
XX Determining if a compound induces a toxic effect on a tissue or cell, for
PT identifying hepatotoxic compounds, comprises comparing a gene expression
PT profile of a tissue or cell sample to a database of Tox mean and non-Tox
PT mean values.
XX
XX Claim 44; SEQ ID NO 4174; 874pp; English.
XX
XX The present invention describes a method for determining whether a
CC compound induces a toxic effect on a tissue or cell. The method comprises
CC preparing a gene expression profile of a tissue or cell sample exposed to
```

```
CC the compound, and comparing the gene expression profile to a database
CC comprising data or information on the Tox mean and non-Tox mean value.
CC The method is useful for predicting or identifying at least one toxic
CC effect, particularly hepatotoxicity, of a test or unknown compound. The
CC genes listed in the specification are useful as diagnostic or toxicity
CC markers for the prediction or identification of the physiological state
CC of tissue or cell sample that has been exposed to a compound, or to
CC identify or predict the toxic effects of a compound or an agent. These
CC may also be used as markers for monitoring toxicity progression or for
CC drug screening. The present sequence represents a primary rat hepatocyte
CC toxicity modelling related gene sequence from the present invention.
XX
SQ Sequence 2543 BP; 725 A; 554 C; 627 G; 637 T; 0 U; 0 Other;
Query Match 41.9%; Score 63.2; DB 9; Length 2543;
Best Local Similarity 74.1%; Pred. No. 1.2e-11;
Matches 80; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
QY 10 CAGGTACTAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGCATGTTGTA 69
DB 1535 CAAGTTCCTAACAGGCCAGATGCTGCTGAGGCTCTGCTTCAACACATAGTGTGG 1476
QY 70 AGCGCGCATCAACCTTCTCTGCTATTATAAATAAAATGTCATAAACT 117
DB 1475 AGCGCTGATCAACCTTCTCTGCTATCACTTGCAGGTGGCAAACCTCT 1428
RESULT 13
ABL55204/c
ID ABL55204 standard; DNA; 3215 BP.
XX
AC ABL55204;
XX
DT 14-JUN-2002 (first entry)
DE HBV subtype adr surface antigen (HBsAg)/core antigen encoding DNA.
XX
KW HBV; subtype adr; surface antigen; HBsAg; core antigen; detection;
KW hepatitis B; gene; ds.
XX
OS Hepatitis B virus.
PH Key Location/Qualifiers
FT primer_bind /tag= a complement(15..34)
FT CDS /note= "Primer S-N1/I (ABL55201)"
FT 29..709
FT /tag= b
FT /product= "Surface antigen (HBsAg; AAM49180)"
FT primer_bind complement(60..79)
FT /tag= c
FT /note= "Primer S-02/I (ABL55207)"
FT primer_bind complement(70..89)
FT /tag= d
FT /note= "Primer S-03/I (ABL55208)"
FT misc_binding complement(150..230)
FT /tag= e
FT /bound moiety= "HBV capture probe (ABL55199)"
FT 221..250
FT /tag= f
FT /bound moiety= "ASCap7-Inosine probe (ABL55205)"
FT 698..723
FT primer_bind /tag= g
FT /note= "Primer AS-2/I (ABL55203)"
FT 710..733
FT primer_bind /tag= h
FT /note= "Primer AS-3/I (ABL55202)"
FT 721..745
FT primer_bind /tag= i
FT /note= "Primer AS-4/I (ABL55209)"
FT 1775..2336
FT CDS /tag= j
FT /product= "Core antigen (AAM49181)"
```

```

FT misc_feature 2780..3215
FT /*tag= k
FT /note= "Corresponds to bases 1-436 of ABL55200"
XX
PN JP2001352989-A.
XX
XX 25-DEC-2001.
XX
XX 14-JUN-2000; 2000JP-00178470.
XX
XX 14-JUN-2000; 2000JP-00178470.
XX
XX (FURE ) FUJIREBIO KK.
XX
XX WPI; 2002-191858/25.
XX
XX P-PSDB; AAM49180, AAM49181.
XX
XX Novel DNA used as a capturing probe for detecting hepatitis B virus.
XX
XX Example 1; Page 14-16; 21pp; Japanese.
XX
XX The invention relates to nucleic acid sequences used to detect hepatitis
XX B virus (HBV). The invention specifically claims a sense capture probe
XX (ABL55199) corresponding to a fragment of the surface antigen (HBsAg)
XX coding sequence (bases 150-290 of the HBV subtype adr DNA shown in
XX ABL55204), and to inosine-containing primers (ABL55201-ABL55202) which
XX bind to regions of the HBV genome in or near the HBsAg coding sequence.
XX The invention also relates to DNA molecules in which at least one base is
XX deleted, replaced or added in the capture probe sequence, and which
XX hybridises to HBV DNA between bases 21-710 of ABL55204, and additionally
XX encompasses methods and reagents used in detecting HBV using the capture
XX probe of the invention. The present sequence represents a fragment of the
XX HBV subtype adr genome encoding the surface and core antigens used in an
XX exemplification of the invention
XX
XX Sequence 3215 BP; 717 A; 860 C; 719 G; 919 T; 0 U; 0 Other;

Query Match 22.1%; Score 33.4; DB 6; Length 3215;
Best Local Similarity 55.7%; Pred. No. 0.46;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 18 TAACAGGCCAAATGTCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGAAGCGCGC 77
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1679 TGAACAGACCAATTTATGCTTACAGCCCTCTAGTACAAAGACCACTTAATCTCTC 1620
QY 78 ATCAACCTCTCTGCTATTAAATAAATGCTATTAACCTCATCTGCAAGGTG 132
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1619 CCCCACTCTCCAGCTTTTAAACAAACAGTCTTTGAATATGCTCAAGGTG 1565

RESULT 14
AAF22296
ID AAF22296 standard; DNA; 129021 BP.
XX
XX AAF22296;
XX
XX 20-MAR-2001 (first entry)
XX
XX BAC containing repeats from centromeres 1-4 #19.
XX
XX Centromere; microsome; vector; ds.
XX
XX Arabidopsis thaliana.
XX
XX WO200055325-A2.
XX
XX 21-SEP-2000.
XX
XX 17-MAR-2000; 2000WO-US007392.
XX
XX 18-MAR-1999; 99US-0125219P.
XX
XX 01-APR-1999; 99US-0127409P.
XX
XX 18-MAY-1999; 99US-0134770P.

PR 13-SEP-1999; 99US-0153584P.
PR 17-SEP-1999; 99US-0154603P.
PR 16-DEC-1999; 99US-0172493P.
XX
XX (UYCH-) UNIV CHICAGO.
XX
XX Preuss D, Copenhaver G, Keith K;
XX
XX WPI; 2000-587529/55.
XX
XX Recombinant DNA construct comprising a plant centromere, useful for
XX producing stably inherited microchromosomes which can serve as vectors for the
XX construction of transgenic plant and animal cells.
XX
XX Claim 102; Page 686-716; 1449pp; English.
XX
XX The present invention relates to a recombinant DNA construct of a plant
XX (Arabidopsis thaliana) centromere. The constructs are useful for
XX producing stably inherited microchromosomes which can serve as vectors for the
XX construction of transgenic plant and animal cells expressing selected
XX proteins such as hormones, enzymes, interleukins, clotting factors,
XX cytokines, antibodies, and growth factors
XX
XX Sequence 129021 BP; 42091 A; 22610 C; 24583 G; 39735 T; 0 U; 2 Other;

Query Match 21.3%; Score 32.2; DB 3; Length 129021;
Best Local Similarity 59.1%; Pred. No. 5.4;
Matches 55; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 27 CAATGTTGGCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGAAGCGCGCATCAACCTT 86
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
124933 CTAGTGCTCTTTCAACTTTCAGTTGAAAGATCGTGTACAGGAGACTGTAAACACCTT 124992
QY 87 CTCGCTCTATTAAATAAATGTCTATAAATCA 119
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
124993 CCATGCCGAGCAGAATCAATGTCTAAGCTGA 125025

RESULT 15
ABT23617/c
ID ABT23617 standard; DNA; 560 BP.
XX
XX AC ABT23617;
XX
XX 22-MAY-2003 (first entry)
XX
XX Stabilising reagent method related oligo SEQ ID No 69.
XX
XX Stabilising reaction reagent; PCR; primer; RNaseH; long-term storage;
XX specific amplification; pathogenic microorganism; chimeric;
XX genetic engineering; clinical medicine; ss.
XX
XX Hepatitis B virus.
XX
XX WO2002101042-A1.
XX
XX 19-DEC-2002.
XX
XX 12-JUN-2002; 2002WO-JP005832.
XX
XX 12-JUN-2001; 2001JP-00177737.
XX
XX 20-AUG-2001; 2001JP-00249689.
XX
XX (TAKI ) TAKARA BIO INC.
XX
XX Sagawa H, Uemori T, Mukai H, Yamamoto J, Tomono J, Kobayashi E;
XX Enoki T, Asada K, Kato I;
XX
XX WPI; 2003-148805/14.
XX
XX Method for stabilizing and storing reaction reagents for specific
XX amplification and detection of nucleic acids particularly in e.g.
XX identifying pathogenic microorganisms or viruses in sample.

```

```

XX
PS
XX
XX
CC The invention relates to a novel stabilising reaction reagent for use in
CC the amplification and/or detection of a target nucleic acid comprising:
CC preparing a reaction mixture with e.g. a nucleic acid as template, at
CC least 1 primer and RNaseH; and incubation of the reaction mixture for a
CC defined period of time to form a reaction product during the
CC amplification of such target nucleic acid. The method is useful for
CC stabilising and long-term storage of reaction reagents for highly
CC sensitive and specific amplification and detection of nucleic acids
CC particularly in identifying pathogenic microorganisms or viruses in a
CC sample using chimeric oligonucleotide primers, which is useful in genetic
CC engineering and clinical medicine. This polynucleotide sequence
CC represents an oligo relating to the novel stabilising reaction reagent
CC method of the invention
XX
SQ Sequence 560 BP; 101 A; 167 C; 143 G; 149 T; 0 U; 0 Other;
Query Match 21.1%; Score 31.8; DB 7; Length 560;
Best Local Similarity 54.8%; Pred. No. 0.84;
Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
Qy 18 TAAACAGGCCAAATGTTGCCCTTTGGGGTTCTCTTTCAACAGCATGGTGAAGCGCGC 77
Db 439 TGAACAGACCAATTTATGCCCTCCTAGCAGCTCTAGTACAAAGACCTTTAACTATCTCTC 380
Qy 78 ATCAACCTTCTGCGCTATTAAATATAATGATCATAACTCATCTGCAAGGTGG 132
Db 379 CCCCACTCTCCAGTCTTTAAACAAACAGTCTTTGAAGTATGCCCTCAAGGTGCG 325

```

Search completed: August 10, 2004, 21:54:08
Job time : 64.5488 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 10, 2004, 21:58:32 ; Search time 70.0822 Seconds
(without alignments)
10571.835 Million cell updates/sec

Title: US-10-001-857-41

Perfect score: 151

Sequence: 1 ccgccggcgaggtacctaa.....gcaaatctctcaagaatatg 151

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3225727 seqs, 2453303834 residues

Total number of hits satisfying chosen parameters: 6451454

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	151	100.0	151	14	US-10-001-857-41
c	140.4	93.0	3096	14	Sequence 41, Appl
3	93.8	62.1	602	10	Sequence 42, Appl
4	93.8	62.1	602	10	Sequence 3218, Ap
5	93.8	62.1	745	10	Sequence 9543, Ap
c	84.8	56.2	2488	17	Sequence 15927, A
7	83.8	55.5	815	14	Sequence 111, App
8	83.2	55.1	2610	10	Sequence 20725, A
c	33.4	22.1	3215	13	Sequence 296, App
c	33.4	22.1	3215	13	Sequence 297, App
c	31.8	21.1	3215	13	Sequence 298, App
c	31.8	21.1	3215	13	Sequence 290, App
c	31.8	21.1	3215	13	Sequence 291, App
c	31.8	21.1	3215	13	Sequence 292, App

c	15	31.8	21.1	3215	13	US-10-453-792-295	Sequence 295, App
c	16	31.8	21.1	3215	13	US-10-443-801A-11	Sequence 11, Appl
c	17	31.4	20.8	3213	13	US-10-453-792-288	Sequence 288, App
c	18	31.4	20.8	3213	13	US-10-453-792-289	Sequence 289, App
c	19	31	20.5	1644	13	US-10-282-122A-23451	Sequence 23451, A
c	20	31	20.5	3161	13	US-10-453-792-301	Sequence 301, App
c	21	30.8	20.4	582	13	US-10-027-632-221960	Sequence 221960,
c	22	30.8	20.4	582	16	US-10-027-632-221960	Sequence 221960,
c	23	30.6	20.3	322101	13	US-10-354-247-1	Sequence 1, Appli
c	24	30.6	20.3	322101	15	US-10-060-902-1	Sequence 1, Appli
c	25	30.2	20.0	711	13	US-10-027-632-149134	Sequence 149134,
c	26	30.2	20.0	711	16	US-10-027-632-149134	Sequence 149134,
c	27	30.2	20.0	3214	13	US-10-453-792-294	Sequence 294, App
c	28	30.2	20.0	3215	13	US-10-453-792-286	Sequence 286, App
c	29	30.2	20.0	197526	17	US-10-322-281-498	Sequence 498, App
c	30	30	19.9	3480	13	US-10-027-632-114041	Sequence 114041,
c	31	30	19.9	3480	13	US-10-027-632-114042	Sequence 114042,
c	32	30	19.9	3480	16	US-10-027-632-114041	Sequence 114041,
c	33	30	19.9	3480	16	US-10-027-632-114042	Sequence 114042,
c	34	30	19.9	13273	9	US-09-764-869-2349	Sequence 2349, Ap
c	35	30	19.9	13273	15	US-10-091-504-2349	Sequence 2349, Ap
c	36	30	19.9	13273	16	US-10-227-577-2349	Sequence 2349, Ap
c	37	30	19.9	300000	15	US-10-262-552-33	Sequence 33, Appl
c	38	30	19.9	300000	17	US-10-703-210-33	Sequence 33, Appl
c	39	29.8	19.7	1510	13	US-10-424-599-125454	Sequence 125454,
c	40	29.6	19.6	576	13	US-10-282-122A-29360	Sequence 29360, A
c	41	29.6	19.6	1497	16	US-10-369-493-33434	Sequence 33434, A
c	42	29.6	19.6	4408	17	US-10-437-963-56946	Sequence 56946, A
c	43	29.4	19.5	197	9	US-09-837-297-2	Sequence 2, Appli
c	44	29.4	19.5	250	10	US-09-875-453-16	Sequence 16, Appl
c	45	29.4	19.5	258	13	US-10-424-599-63730	Sequence 63730, A

ALIGNMENTS

RESULT 1

US-10-001-857-41
; Sequence 41, Application US/10001857
; Publication No. US20020183500A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Pro
; FILE REFERENCE: DEX-0273
; CURRENT APPLICATION NUMBER: US/10/001,857
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,054
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41
; LENGTH: 151
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-001-857-41

Query Match 100.0%; Score 151; DB 14; Length 151;
Best Local Similarity 100.0%; Pred. No. 3.1e-42;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	CCGCCCGGCGAGGTACTAAACAGGCCAAATGTTGGCTTTGGGGTTCCTGTTTCAACAGC	60
Db	1	CCGCCCGGCGAGGTACTAAACAGGCCAAATGTTGGCTTTGGGGTTCCTGTTTCAACAGC	60
Qy	61	ATGGTGTGAAGCGCGCGATCAACCTTCTGTGCTTATAAATAAATGTCATTAACCTCAT	120
Db	61	ATGGTGTGAAGCGCGCGATCAACCTTCTGTGCTTATAAATAAATGTCATTAACCTCAT	120
Qy	121	CCTGCAAGGTGGCAAAATTCCTCAAGATATG	151

```
Db 121 CCTGCAAGTGGCAAAATTCCTCAAGAAATG 151
|||||
RESULT 2
US-10-001-857-42/c
; Sequence 42, Application US/10001857
; Publication No. US20020183500A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Hervé
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Proteins
; FILE REFERENCE: DEX-0273
; CURRENT APPLICATION NUMBER: US/10/001,857
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,054
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
; LENGTH: 3096
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-001-857-42
Query Match 93.0%; Score 140.4; DB 14; Length 3096;
Best Local Similarity 99.3%; Pred. No. 6.4e-38;
Matches 141; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 CAGTACTTAACAGGCAAAATGTCCTTTGGGGTTCCTTTCAACAGCATGGTGGA 69
Db 2052 CAGTACTTAACAGGCAAAATGTCCTTTGGGGTTCCTTTCAACAGCATGGTGGA 1993
|||||
Qy 70 AGCCCGCATCAACCTTCTGCTTATTAATAAATGTCATAAATCACTTCCTGCAAGG 129
Db 1992 AGCGTGCATCAACCTTCTGCTTATTAATAAATGTCATAAATCACTTCCTGCAAGG 1993
|||||
Qy 130 TGGCAAAATTCCTCAAGAAATG 151
Db 1932 TGGCAAAATTCCTCAAGAAATG 1911
|||||
RESULT 3
US-09-814-353-3218
; Sequence 3218, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9543
; LENGTH: 602
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 561..590
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-3218
Query Match 62.1%; Score 93.8; DB 10; Length 602;
Best Local Similarity 89.4%; Pred. No. 4.8e-22;
Matches 101; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 CCGCCCGGCGAGGTACCTTAACAGGCCAAATGTTGCCCTTTGGGGTTCCTGTTTCAACAGC 60
Db 14 CCGCCCGGCGAGGTACCTTAACAGGCCAAATGTTGCCCTTTGGGGTTCCTGTTTCAACAGC 73
|||||
Qy 61 ATGGTGTGAAGCGCGCATCAACCTTCTGCTTATTAATAAATGTCATA 113
Db 74 ATGGTGTGAAGCGCTGCATCAACCTTCTGCTTATTAATAAATGTCATA 126
|||||
RESULT 4
US-09-814-353-9543
; Sequence 9543, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9543
; LENGTH: 602
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 561..590
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-9543
Query Match 62.1%; Score 93.8; DB 10; Length 602;
Best Local Similarity 89.4%; Pred. No. 4.8e-22;
Matches 101; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 CCGCCCGGCGAGGTACCTTAACAGGCCAAATGTTGCCCTTTGGGGTTCCTGTTTCAACAGC 60
Db 14 CCGCCCGGCGAGGTACCTTAACAGGCCAAATGTTGCCCTTTGGGGTTCCTGTTTCAACAGC 73
|||||
Qy 61 ATGGTGTGAAGCGCGCATCAACCTTCTGCTTATTAATAAATGTCATA 113
Db 74 ATGGTGTGAAGCGCTGCATCAACCTTCTGCTTATTAATAAATGTCATA 126
|||||
```

```
RESULT 5
US-09-814-353-15927
; Sequence 15927, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15927
; LENGTH: 745
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 571
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-15927

Query Match      62.1%; Score 93.8; DB 10; Length 745;
Best Local Similarity 89.4%; Pred. No. 5.3e-22;
Matches 101; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 CCGCCGGCAGGTACTTAACAGGCCAAATGTCCTTGGGTTCTCTTTTCAACAGC 60
Db 101 CCGCCGGCAGGTACTTAACAGGCCAAATGTCCTTGGGTTCTCTTTTCAACAGC 160
QY 61 ATGGTGTGAAGCGCGCATCAACCTTCTCTGCTATTAAAAATAAAGTCATA 113
Db 161 ATGGTGTGAAGCGCGCATCAACCTTCTCTGCTATTAAAAATAAAGTCATA 213

RESULT 6
US-10-467-433-23/c
; Sequence 23, Application US/10467433
; Publication No. US2004008773A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION;
; APPLICANT: LAL, Preeti G.; BAUGHN, Mariah R.;
; APPLICANT: YAO, Monique G.; CHAWLA, Narinder K.;
; APPLICANT: ELLIOTT, Vicki S.; XU, Yuming;
; APPLICANT: HONCHELL, Cynthia D.; YUE, Henry;
; APPLICANT: DING, Li; GIETZEN, Kimberly J.;
; APPLICANT: ISON, Craig H.; LU, Dying Aina M.;
; APPLICANT: HAFALIA, April J.A.; SANDHI, Ameena R.;
; APPLICANT: THANGAVELU, Kavitha; SANJANWALA, Madhusudan M.;
; APPLICANT: TANG, Y. Tom; RAMKUMAR, Jayalaxmi;
; APPLICANT: GRIFFIN, Jennifer A.; SWARNAKAR, Anita;
; APPLICANT: AZIMZAI, Yalda; SAPPERSTEIN, Stephanie K.;
; APPLICANT: BURFORD, Neil; LEE, Ernestine A.;
; APPLICANT: LU, Yan; TRAN, Uyen K.;
; APPLICANT: MARQUIS, Joseph P.
; TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT
```

```
FILE REFERENCE: PF-0899 USN
; CURRENT APPLICATION NUMBER: US/10/467,433
; CURRENT FILING DATE: 2003-08-06
; PRIOR APPLICATION NUMBER: PCT/US02/03709
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/268,117
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: US 60/269,618
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: US 60/271,118
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: US 60/274,486
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 60/274,436
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/334,229
; PRIOR FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: US 60/353,284
; PRIOR FILING DATE: 2002-02-01
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PERL Program
; SEQ ID NO 23
; LENGTH: 2488
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 7111920CBI
US-10-467-433-23

Query Match      56.2%; Score 84.8; DB 17; Length 2488;
Best Local Similarity 88.5%; Pred. No. 1.3e-18;
Matches 92; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 10 CAGGTACCTAAACAGGCCAAATGTCCTTGGGGTTCTCTTCAACAGCATGGTGA 69
Db 1554 CAGGTACCTAAACAGGCCAAATGTCCTTGGGGTTCTCTTCAACAGCATGGTGA 1495
QY 70 AGGCGCGCATCAACCTTCTCTGCTATTAAAAATAAATGTCATA 113
Db 1494 AGGCGCGCATCAACCTTCTCTGCTATTAAAAATAAATGTCATA 1451

RESULT 7
US-10-001-857-111
; Sequence 111, Application US/10001857
; Publication No. US20020183500A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prot
; FILE REFERENCE: DEX-0273
; CURRENT APPLICATION NUMBER: US/10/001,857
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,054
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 111
; LENGTH: 815
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (206)..(305)
; OTHER INFORMATION: a, c, g or t
US-10-001-857-111

Query Match      55.5%; Score 83.8; DB 14; Length 815;
Best Local Similarity 88.3%; Pred. No. 1.7e-18;
```

Matches 91; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 11 AGGTACCTAAACAGCCAAATGTCCTTGGGGTTCCTTTCAACAGCATGGTGCA 70
Db 87 AGGTACCTAAACAGCCAAATGTCCTTGGGGTTCCTTTCAACAGCATGGTGCA 146
QY 71 GCGCGCATCAACCTTCTCGCTATTAAATAAAATGTCATA 113
Db 147 GCGCTGCATCAACCTTCTCGCTCATCTGCAAGGTGGCAA 189

RESULT 8

US-09-814-353-20725
; Sequence 20725, Application US/09814353
; Publication No. US20030165931A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20725
; LENGTH: 2610
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1, 2, 3, 4
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-20725

Query Match 55.1%; Score 83.2; DB 10; Length 2610;
Best Local Similarity 87.5%; Pred. No. 4.7e-18;
Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY 10 CAGGTACCTAAACAGCCAAATGTCCTTGGGGTTCCTTTCAACAGCATGGTGCA 69
Db 1009 CAGGTACCTAAACAGCCAAATGTCCTTGGGGTTCCTTTCAACAGCATGGTGCA 1068
QY 70 AGCGCGCATCAACCTTCTCGCTATTAAATAAAATGTCATA 113
Db 1069 AGCGTGCATCAACCTTCTCGCTCATCTGCAAGGTGGCAA 1112

RESULT 9

US-10-453-792-296/c
; Sequence 296, Application US/10453792
; Publication No. US20040029110A1
; GENERAL INFORMATION:
; APPLICANT: STUYVER, LIEVEN
; APPLICANT: ROSSAU, RUDI
; APPLICANT: MAERTENS, GEERT
; TITLE OF INVENTION: METHOD FOR TYPING AND DETECTING HBV
; NUMBER OF SEQUENCES: 313
; CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/453,792
FILING DATE: 04-Jun-2003
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/155,885A
FILING DATE: 08-Oct-1998
APPLICATION NUMBER: PCT/EP97/02002
FILING DATE: 21-Apr-1997
APPLICATION NUMBER: EP 96870053.4
FILING DATE: 19-Apr-1996
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2551-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 296:
SEQUENCE CHARACTERISTICS:
LENGTH: 3215 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 296:
US-10-453-792-296

Query Match 22.1%; Score 33.4; DB 13; Length 3215;
Best Local Similarity 55.7%; Pred. No. 1.1;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
QY 18 TAAACAGCCAAATGTCCTTGGGGTTCCTTTCAACAGCATGGTGCAAGCGCGC 77
Db 1807 TGAACAGACCAATTATGCTTACAGCCTCTAGTACAAAGACCATTAACCTAATCTCTC 1748
QY 78 ATCAACCTTCTCGCTATTAAATAAAATGTCATAAACTCATCTGCAAGTGG 132
Db 1747 CCCCACTCTCCAGCTTTTAAACAAACAGCTTTGAATATGCTCAAGTGC 1693

RESULT 10

US-10-453-792-297/c
; Sequence 297, Application US/10453792
; Publication No. US20040029110A1
; GENERAL INFORMATION:
; APPLICANT: STUYVER, LIEVEN
; APPLICANT: ROSSAU, RUDI
; APPLICANT: MAERTENS, GEERT
; TITLE OF INVENTION: METHOD FOR TYPING AND DETECTING HBV
; NUMBER OF SEQUENCES: 313
; CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk


```
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/453,792
; FILING DATE: 04-Jun-2003
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/155,885A
; FILING DATE: 08-Oct-1998
; APPLICATION NUMBER: PCT/EP97/02002
; FILING DATE: 21-APR-1997
; APPLICATION NUMBER: EP 96870053.4
; FILING DATE: 19-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 2551-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 297:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3215 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 297:
US-10-453-792-297

Query Match 22.1%; Score 33.4; DB 13; Length 3215;
Best Local Similarity 55.7%; Pred. No. 1.1;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 18 TAAACAGGCCAAATGTTGGCTTCTCTTCAACAGCATGTTGTAAGCGCGC 77
Db 1807 TGAACAGACCAATTTAGCTCCTAGTACAAAGACCATTAACCTATCTCCTC 1748

QY 78 ATCAACCTTCTCTGCCTATTAAATAAATGTATCAAACTCATCTCCTCAAGGTGG 132
Db 1747 CCCCACTCTCCAGTCTTTAAACAACAGCTCTTTGAATATGCTCAAGGTG 1693

RESULT 11
US-10-453-792-298/c
; Sequence 298, Application US/10453792
; Publication No. US20040029110A1
; GENERAL INFORMATION:
; APPLICANT: STUYVER, LIEVEN
; ROSSAU, RUDI
; MAERTENS, GEERT
; TITLE OF INVENTION: METHOD FOR TYPING AND DETECTING HBV
; NUMBER OF SEQUENCES: 313
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/453,792
; FILING DATE: 04-Jun-2003
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
```

```
;
; APPLICATION NUMBER: US/09/155,885A
; FILING DATE: 08-Oct-1998
; APPLICATION NUMBER: PCT/EP97/02002
; FILING DATE: 21-APR-1997
; APPLICATION NUMBER: EP 96870053.4
; FILING DATE: 19-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 2551-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 298:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3212 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 298:
US-10-453-792-298

Query Match 21.1%; Score 31.8; DB 13; Length 3212;
Best Local Similarity 54.8%; Pred. No. 4;
Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 18 TAAACAGGCCAAATGTTGGCTTCTCTTCAACAGCATGTTGTAAGCGCGC 77
Db 1804 TGAACAGACCAATTTAGCTCCTAGTACAAAGATCATTAACCTAAGTTCCTC 1745

QY 78 ATCAACCTTCTCTGCCTATTAAATAAATGTATCAAACTCATCTCCTCAAGGTGG 132
Db 1744 CCCCACTCTCCAGTCTTTAAACAACAGCTCTTTGAATATGCTCAAGGTG 1690

RESULT 12
US-10-453-792-290/c
; Sequence 290, Application US/10453792
; Publication No. US20040029110A1
; GENERAL INFORMATION:
; APPLICANT: STUYVER, LIEVEN
; ROSSAU, RUDI
; MAERTENS, GEERT
; TITLE OF INVENTION: METHOD FOR TYPING AND DETECTING HBV
; NUMBER OF SEQUENCES: 313
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/453,792
; FILING DATE: 04-Jun-2003
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/155,885A
; FILING DATE: 08-Oct-1998
; APPLICATION NUMBER: PCT/EP97/02002
; FILING DATE: 21-APR-1997
; APPLICATION NUMBER: EP 96870053.4
; FILING DATE: 19-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
```

REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2551-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 290:
SEQUENCE CHARACTERISTICS:
LENGTH: 3215 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 290:

US-10-453-792-290
Query Match 21.1%; Score 31.8; DB 13; Length 3215;
Best Local Similarity 54.8%; Pred. No. 4;
Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
QY 18 TAAACAGCCCAATGTCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGAAGCGCGC 77
DB 1807 TGAACAGACCAATTTATGCTTACAGCCTCTAGTACAAAGACCTTTAACTTAATCTCTC 1748
QY 78 ATCAACCTTCTGCTTATTAATAAATAATGCTATCAAACTCATCTGCAAGGTGG 132
DB 1747 CCCCAACTCTCCAGTCTTTAAACAAACAGCTTTGAAGTATGCTCAAGTCG 1693

RESULT 13

US-10-453-792-291/c
Sequence 291, Application US/10453792
Publication No. US20040029110A1
GENERAL INFORMATION:
APPLICANT: STUYVER, LIEVEN
ROSSAU, RUDI
MAERTENS, GEERT
TITLE OF INVENTION: METHOD FOR TYPING AND DETECTING HBV
NUMBER OF SEQUENCES: 313
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/453,792
FILING DATE: 04-Jun-2003
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/155,885A
FILING DATE: 08-Oct-1998
APPLICATION NUMBER: PCT/EP97/02002
FILING DATE: 21-APR-1997
APPLICATION NUMBER: EP 96870053.4
FILING DATE: 19-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2551-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 291:
SEQUENCE CHARACTERISTICS:
LENGTH: 3215 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 291:

US-10-453-792-291
Query Match 21.1%; Score 31.8; DB 13; Length 3215;
Best Local Similarity 54.8%; Pred. No. 4;
Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
QY 18 TAAACAGCCCAATGTCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGAAGCGCGC 77
DB 1807 TGAACAGACCAATTTATGCTTACAGCCTCTAGTACAAAGACCTTTAACTTAATCTCTC 1748
QY 78 ATCAACCTTCTGCTTATTAATAAATAATGCTATCAAACTCATCTGCAAGGTGG 132
DB 1747 CCCCAACTCTCCAGTCTTTAAACAAACAGCTTTGAAGTATGCTCAAGTCG 1693

RESULT 14

US-10-453-792-292/c
Sequence 292, Application US/10453792
Publication No. US20040029110A1
GENERAL INFORMATION:
APPLICANT: STUYVER, LIEVEN
ROSSAU, RUDI
MAERTENS, GEERT
TITLE OF INVENTION: METHOD FOR TYPING AND DETECTING HBV
NUMBER OF SEQUENCES: 313
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/453,792
FILING DATE: 04-Jun-2003
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/155,885A
FILING DATE: 08-Oct-1998
APPLICATION NUMBER: PCT/EP97/02002
FILING DATE: 21-APR-1997
APPLICATION NUMBER: EP 96870053.4
FILING DATE: 19-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2551-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 292:
SEQUENCE CHARACTERISTICS:
LENGTH: 3215 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 292:

US-10-453-792-292

```

Query Match          21.1%; Score 31.8; DB 13; Length 3215;
Best Local Similarity 54.8%; Pred. No. 4;
Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY      18  TAAACAGCGCCAAATGTTGCCCTTTGGGTTTCCTGTTTCAACAGCATGTTGGAAGCGCGC 77
Db      1807  TGAACAGACCAATTTAGCTCTACAGCCTCTCTAGTACAAAGACCTTTTAACCTAATCTCTC 1748

QY      78  ATCAACCTTCTCTGCGCTATTAAAAATAAAATGTCATAACTCATCTCTGCAAGGTGG 132
Db      1747  CCCCAACTCTCCAGTCTTTTAAACAACAGTCTTTGAAGTATGCTTCAAGTCC 1693

RESULT 15
US-10-453-792-295/c
; Sequence 295, Application US/10453792
; Publication No. US20040029110A1
; GENERAL INFORMATION:
; APPLICANT: STUYVER, LIEVEN
; ROSSAU, RUDI
; MAERTENS, GEERT
; TITLE OF INVENTION: METHOD FOR TYPING AND DETECTING HBV
; NUMBER OF SEQUENCES: 313
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/453,792
; FILING DATE: 04-Jun-2003
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/155,885A
; FILING DATE: 08-Oct-1998
; APPLICATION NUMBER: PCT/EP97/02002
; FILING DATE: 21-APR-1997
; APPLICATION NUMBER: EP 96870053.4
; FILING DATE: 19-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 2551-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 295:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3215 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 295:
US-10-453-792-295

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: August 10, 2004, 17:10:22 ; Search time 12336.3 Seconds
(without alignments)
10877.634 Million cell updates/sec

Title: US-10-001-857-42
Perfect score: 3096
Sequence: 1 ttctcacgaactccagg.....attcttgacaaaaaaaaa 3096

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	3096	100.0	3096	6	AX523454	AX523454 Sequence
2	2417.2	78.1	2488	6	AX537452	AX537452 Sequence
3	1838.4	59.4	2552	10	AY102701	AY102701 Mus muscu
4	1828.2	59.1	2594	10	BC056435	BC056435 Mus muscu
5	1799.2	58.1	2543	10	AF272892	AF272892 Rattus no
6	1766.6	57.1	1839	9	AK025266	AK025266 Homo sapi
7	1264.8	40.9	2467	5	BC053286	BC053286 Danio rer
8	988	31.9	997	9	AK026296	AK026296 Homo sapi
9	885.4	28.6	1796	9	AK056059	AK056059 Homo sapi
10	717.4	23.2	719	6	AK664135	AK664135 Sequence
11	664.2	21.5	147876	2	AC102010	AC102010 Mus muscu
12	551.4	17.8	181864	9	AL353743	AL353743 Human DNA
13	440	14.2	440	6	AX778899	AX778899 Sequence
14	398.4	12.9	160115	9	AL161447	AL161447 Human DNA
C 15	321.4	10.4	446	9	HSR334821	HSR334821 Homo sapi
C 16	299	9.7	743	9	HSR334413	HSR334413 Homo sapi
17	236.4	7.6	642	9	HSR342328	HSR342328 Homo sapi
18	236.4	7.6	742	9	HSR331811	HSR331811 Homo sapi
19	233.2	7.5	640	9	HSR342329	HSR342329 Homo sapi
20	233.2	7.5	652	9	HSR342327	HSR342327 Homo sapi
21	217	7.0	2537	3	AY060948	AY060948 Drosophil
22	207.6	6.7	2565	3	AY089452	AY089452 Drosophil
23	207	6.7	320902	2	AL161456	AL161456 Homo sapi
C 24	182	5.9	489	11	G63197	G63197 SHGC-140830
25	177.8	5.7	34935	9	AL161453	AL161453 Human DNA
26	177.6	5.7	111554	2	AC143812	AC143812 Macaca mu
C 27	155.8	5.0	113836	10	AL669837	AL669837 Mouse DNA
28	153.2	4.9	253611	2	AC111473	AC111473 Rattus no
C 29	153.2	4.9	271861	2	AC108541	AC108541 Rattus no
30	153.2	4.9	281804	2	AC134869	AC134869 Mus muscu
31	147.2	4.8	201890	2	AC135469	AC135469 Mus muscu
C 32	140.4	4.5	151	6	AX523453	AX523453 Sequence
33	134.4	4.3	204131	2	AC128898	AC128898 Rattus no
34	130.4	4.2	3497	10	AY102702	AY102702 Mus muscu
35	90.6	2.9	16016	2	AC020508	AC020508 Drosophil
36	90.6	2.9	60135	2	AC091200	AC091200 Drosophil
C 37	90.6	2.9	164443	3	AC012167	AC012167 Drosophil
C 38	90.6	2.9	298827	3	AE003463	AE003463 Drosophil
C 39	87.8	2.8	815	6	AX523523	AX523523 Sequence
C 40	69.4	2.2	7218	6	I66494	I66494 Sequence 14
41	66.6	2.2	2000	6	AX655393	AX655393 Sequence
C 42	66.6	2.2	110737	2	AC011105	AC011105 Homo sapi
C 43	65.6	2.1	68328	2	AC137705	AC137705 Homo sapi
44	64.6	2.1	832	11	BV026280	BV026280 S212P647
C 45	64.4	2.1	69373	2	AC101456	AC101456 Mus muscu

ALIGNMENTS

RESULT 1
AX523454
LOCUS AX523454 3096 bp DNA linear PAT 24-OCT-2002
DEFINITION Sequence 42 from Patent WO02064788.
ACCESSION AX523454
VERSION AX523454.1 GI:24412350
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1
AUTHORS Macina, R.A., Recipon, H., Chen, S.Y., Sun, Y. and Liu, C.
TITLE Compositions and methods relating to lung specific genes and proteins


```
Db 1981 TGATGAGCGCTTCAACCATGCTGTGAAACAGAAACCCCAAGGCAACATTTGGCCGTG 2040
QY 2041 TTTAGGTACCTGGGTCTCTTACCAATTAACCTTCGCAATTATGATACAGTACCTCTTAAGTGG 2100
Db 2041 TTTAGGTACCTGGGTCTCTTACCAATTAACCTTCGCAATTATGATACAGTACCTCTTAAGTGG 2100
QY 2101 CTTTGAATTGGAACTCTACAGTATGACAGAGTACTATTATCATATATTTGGTATCTCTCTGA 2160
Db 2101 CTTTGAATTGGAACTCTACAGTATGACAGAGTACTATTATCATATATTTGGTATCTCTCTGA 2160
QY 2161 ATTCCCTTTAGCATGTTGATGTCAACATTTAGTGTGCGGATGGCTCTCAATATGCGAGA 2220
Db 2161 ATTCCCTTTAGCATGTTGATGTCAACATTTAGTGTGCGGATGGCTCTCAATATGCGAGA 2220
QY 2221 GGAAGGATAATGAAAGCAGCAGAGAAAGCCGTAGTATGAAAGAAACAAAGAAAGAAAAA 2280
Db 2221 GGAAGGATAATGAAAGCAGCAGAGAAAGCCGTAGTATGAAAGAAACAAAGAAAGAAAAA 2280
QY 2281 GAAAGTTGCCCATTTGAGCGGAGAGATCACAATGAGCCAGCATATCAGAACATGTGTGC 2340
Db 2281 GAAAGTTGCCCATTTGAGCGGAGAGATCACAATGAGCCAGCATATCAGAACATGTGTGC 2340
QY 2341 TGGAAATGTTTAAACCATGTAGTACATTTGACATGGACGGCAAGTACGTAACCCGAAGTT 2400
Db 2341 TGGAAATGTTTAAACCATGTAGTACATTTGACATGGACGGCAAGTACGTAACCCGAAGTT 2400
QY 2401 TGAGCTTGATAGTGAACAAAGTTCGGTATGAACACAGGTTTGTCTCCATTTCAACAGTGTGAT 2460
Db 2401 TGAGCTTGATAGTGAACAAAGTTCGGTATGAACACAGGTTTGTCTCCATTTCAACAGTGTGAT 2460
QY 2461 GACCCGCGCGAGTGACACTTACATTCAGTTCAAGGAAATGTCTGACCTCAATAAATATAG 2520
Db 2461 GACCCGCGCGAGTGACACTTACATTCAGTTCAAGGAAATGTCTGACCTCAATAAATATAG 2520
QY 2521 CCTCTCTCTCAGTCTCTGAACTGTATGTGGCAGCTAGTAAAGCACTTTCAACAGGCAAA 2580
Db 2521 CCTCTCTCTCAGTCTCTGAACTGTATGTGGCAGCTAGTAAAGCACTTTCAACAGGCAAA 2580
QY 2581 AATGATATTGGAAATATTCTTAACCCGACCATGAGGTTTAAAGGTTGC 2640
Db 2581 AATGATATTGGAAATATTCTTAACCCGACCATGAGGTTTAAAGGTTGC 2640
QY 2641 CAAACCCAACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAGAGGAATCTAAAGT 2700
Db 2641 CAAACCCAACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAGAGGAATCTAAAGT 2700
QY 2701 TCCTCTGAAATTTGATTTCTCTCTCATATAATTTTCTGTTGTGAACTCTGTTTGAGA 2760
Db 2701 TCCTCTGAAATTTGATTTCTCTCTCATATAATTTTCTGTTGTGAACTCTGTTTGAGA 2760
QY 2761 GAGACTGGGGAGTGGCCATAAAGGGGCGAGTCTTTCTTCAGACCCCACTCTTAGAGGG 2820
Db 2761 GAGACTGGGGAGTGGCCATAAAGGGGCGAGTCTTTCTTCAGACCCCACTCTTAGAGGG 2820
QY 2821 CACATCACAGGCTCCACATCACGGAGAGTGAATGATTTCTTGGTAAACAATCTATTA 2880
Db 2821 CACATCACAGGCTCCACATCACGGAGAGTGAATGATTTCTTGGTAAACAATCTATTA 2880
QY 2881 TAAGGAATACTTTTAGTTTGACAGCCTTATATCACATGAATGAAACTCGCTGTTTAAAG 2940
Db 2881 TAAGGAATACTTTTAGTTTGACAGCCTTATATGACATGAATGAAACTCGCTGTTTAAAG 2940
QY 2941 TGGTTTATTATGTTCCATGAAGAAACTGGTCTTATTGATGCAATTTGATGAAGCTTATAT 3000
Db 2941 TGGTTTATTATGTTCCATGAAGAAACTGGTCTTATTGATGCAATTTGATGAAGCTTATAT 3000
QY 3001 GGTTTTATTACAGATTTAATCAAAATCAATTTTTTATGAATGATTTGAGTGAAGAAATAGTGT 3060
Db 3001 GGTTTTATTACAGATTTAATCAAAATCAATTTTTTATGAATGATTTGAGTGAAGAAATAGTGT 3060
QY 3061 TTATAAGGTTAATAAATTTCTTGACAAAGAAAAA 3096
```

```
Db 3061 TTATAAAGGTTAATAAATTTCTTGACAAAGAAAAA 3096

RESULT 2
AX537452
LOCUS AX537452 2488 bp DNA linear PAT 23-NOV-2002
DEFINITION Sequence 23 from Patent WO02070709.
ACCESSION AX537452
VERSION AX537452.1 GI:25269238
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Lal, P.G., Baughn, M.R., Yao, M.G., Wallia, N.K., Elliot, V.S., Xu, Y.,
Honchell, C.D., Yue, H., Ding, L., Gietzen, K.J., Ison, C.H., Lu, D.A.,
Hafalia, A.J., Ghandi, A.R., Thangavelu, K., Sanjanwala, M.M.,
Tang, Y.F., Ramkumar, J., Griffin, J.A., Swarnaker, A., Azimzai, Y.,
Sapperstein, S.K., Burford, N., Lee, E.A., Lu, Y., Tran, U.K. and
Marquis, J.P.
TITLE Molecules for disease detection and treatment
JOURNAL Patent: WO 02070709-A 23 12-SEP-2002;
INCYTE Genomics, Inc. (US)
FEATURES
source
location/Qualifiers
1..2488
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 7111920CB1"

ORIGIN

Query Match 78.1%; Score 2417.2; DB 6; Length 2488;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 2470; Conservative 0; Mismatches 13; Indels 25; Gaps 2;

QY 474 GCGCGCGCGGAGGCGGCGTTCGTTATTTCCGTGTCGGACAGTGGTGGCGCGGGT 533
Db 1 GCGCGCGCGGAGGCGGCGTTCGTTATTTCCGTGTCGGACAGTGGTGGCGCGGGT 60
QY 534 GACCACGGGAGAGTAGGCATATGTTATGAAAGCTTCTGTAGATGATGACCATTCAGG 593
Db 61 GACCACGGGAGAGTAGGCATATGTTATGAAAGCTTCTGTAGATGATGACCATTCAGG 120
QY 594 ATGGGAGCTCAGTATGCCAGAAAAATGGAGAAAAAGCAATACAAACTGGGTGACATTAC 653
Db 121 ATGGGAGCTCAGTATGCCAGAAAAATGGAGAAAAAGCAATACAAACTGGGTGACATTAC 180
QY 654 CCNAGATTTTGAAGAGCTTTCGAGAATTAAAGTTGGGAGAACTACTTCATGATAGCT 713
Db 181 CCNAGATTTTGAAGAGCTTTCGAGAATTAAAGTTGGGAGAACTACTTCATGATAGCT 240
QY 714 ATTGTGCTTTTGAAGCCATGTCCTATTCGAAATGATGGATCCCAAGATGATGCTGG 773
Db 241 ATTGTGCTTTTGAAGCCATGTCCTATTCGAAATGATGGATCCCAAGATGATGCTGG 300
QY 774 CATGATTGAAACCAAGTTTAAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGG 833
Db 301 CATGATTGAAACCAAGTTTAAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGG 360
QY 834 CACTATTAAATTAAGATCTCACCTTGCTGCTGAACTGATAGGATTTATGATACATGTTT 893
Db 361 CACTATTAAATTAAGATCTCACCTTGCTGCTGAACTGATAGGATTTATGATACATGTTT 420
QY 894 TTGCTGTTTGAACCTGTTAGAGGCCATTTCACTGGCACAGACAGTATTTACGTCCT 953
Db 421 TTGCTGTTTGAACCTGTTAGAGGCCATTTCACTGGCACAGACAGTATTTACGTCCT 480
QY 954 TTACATTCATAATCCAGACTTTTATGAAGATCCTGCTATGAAGGCTTTTGTCTCTGGGAAT 1013
Db 481 TTACATTCATAATCCAGACTTTTATGAAGATCCTGCTATGAAGGCTTTTGTCTCTGGGAAT 540
QY 1014 CTTGAAAAATCTGTGACATTCGAAAGGAAAAAGTAAATAAAGCTGCTGTTTTGAGAGGA 1073
```

Db 541. |||||CTTGAAATCTGTGACATTCGAGGAAAGTAAATAAGCTGCTGTTTTGAGAGGA 600
Qy 1074 AGATTTTCAGTCAATGACTTATGATTTAAATGGCTTAACAGTGTGACAGATCTTCGAGT 1133
Db 601 AGATTTTCAGTCAATGACTTATGATTTAAATGGCTTAACAGTGTGACAGATCTTCGAGT 660
Qy 1134 TACAGGCATGCTAAAGATGTGGAGGATGACATCAAGAGAGAGTAAAGTACTTCGAAG 1193
Db 661 TACAGGCATGCTAAAGATGTGGAGGATGACATCAAGAGAGAGTAAAGTACTTCGAAG 720
Qy 1194 TCGACAAGGAGAGAGATCCAGAAAGTTGAACTAGAACACCAACAAATGTTTAGCAGT 1253
Db 721 TCGACAAGGAGAGAGATCCAGAAAGTTGAACTAGAAACCAACAAATGTTTAGCAGT 780
Qy 1254 ATTACAGAGAGTAAATTTACTCGTGTGTTACTGACAGTCTTATAGCTTTTACTAAGAA 1313
Db 781 ATTACAGAGAGTAAATTTACTCGTGTGTTACTGACAGTCTTATAGCTTTTACTAAGAA 840
Qy 1314 AGACACAGTCTGTTGCAGAGCTCAAAATTTGATGGTTTCAAGCAGCAGATCTTCTTTC 1373
Db 841 AGACACAGTCTGTTGCAGAGCTCAAAATTTGATGGTTTCAAGCAGCAGATCTTCTTTC 900
Qy 1374 TGCCATTCAATAATTCATTGCAATCATGGCATCCAGGCCAGAAATGATCTACAAAAGGAGA 1433
Db 901 TGCCATTCAATAATTCATTGCAATCATGGCATCCAGGCCAGAAATGATCTACAAAAGGAGA 960
Qy 1434 TCATCCAAATTAATGATGGTTTGAACCCCTTGTGAACAGAGGCTACTTCCACCTACCTT 1493
Db 961 TCATCCAAATTAATGATGGTTTGAACCCCTTGTGAACAGAGGCTACTTCCACCTACCTT 1020
Qy 1494 CCCTCGATATCAAAATAATTAAGGAGAAAGTGTGAATTTTGAAGATTAAT 1553
Db 1021 CCCTCGATATCAAAATAATTAAGGAGAAAGTGTGAATTTTGAAGATTAAT 1080
Qy 1554 AGATAGAAATAAACTGCTGTGAGGTTGTGAATTTTAAACAAATTTACATTTGATCTCTGGA 1613
Db 1081 AGATAGAAATAAACTGCTGTGAGGTTGTGAATTTTAAACAAATTTACATTTGATCTCTGGA 1140
Qy 1614 TTTTCTCTGTGAATTTAGTGAACAGTACCATGTGTTCTTTCAAGATCTCTGTTACAAC 1673
Db 1141 TTTTCTCTGTGAATTTAGTGAACAGTACCATGTGTTCTTTCAAGATCTCTGTTACAAC 1200
Qy 1674 CACTTCTCTGTGATACAAAAGGCTTTTGGAACTCATCTCATGAAGACATGGTGAA 1733
Db 1201 CACTTCTCTGTGATACAAAAGGCTTTTGGAACTCATCTCATGAAGACATGGTGAA 1260
Qy 1734 AGATGCACTTCGGTCTTTTGTGATCCTCCGAGTGTCTTCCCAAGTGTCTATAT 1793
Db 1261 AGATGCACTTCGGTCTTTTGTGATCCTCCGAGTGTCTTCCCAAGTGTCTATAT 1318
Qy 1794 AATAATCACAGGCTAAGGACTGTATCGACTCTTTTGTACTACTGTGTTCCGCAATC 1853
Db 1319 AATAATCACAGGCTAAGGACTGTATCGACTCTTTTGTACTACTGTGTTCCGCAATC 1378
Qy 1854 TGTAGTCTTATTCAGATCCATGGACATAAAGGGCTCGACAGAGAGATAAGCTTGGTCA 1913
Db 1379 TGTAGTCTTATTCAGATCCATGGACATAAAGGGCTCGACAGAGAGATAAGCTTGGTCA 1438
Qy 1914 ATTCTTGAGGAATTTGCCACTTCGAGGATGAGTTTATGACATTTTATTTAATGAGCAG 1973
Db 1439 ATTCTTGAGGAATTTGCCACTTCGAGGATG-----AGGCAG 1475
Qy 1974 AGAAGGTTGATGACGCGCTTACACCACTGCTGTTGAACAGGAACCCCAAGGCAACATT 2033
Db 1476 AGAAGGTTGATGACGCGCTTACACCACTGCTGTTGAACAGGAACCCCAAGGCAACATT 1535
Qy 2034 TGGCTGTTTAGGTAACCTGGGCTCTTTTACCAATACCTTCGCATTAATGATACAGTACCTTC 2093
Db 1536 TGGCTGTTTAGGTAACCTGGGCTCTTTTACCAATACCTTCGCATTAATGATACAGTACCTTC 1595
Qy 2094 TAAGTGGCTTTGAATTTGAACTCTACAGTATGCAGGACTATATGATATATTGGTATC 2153
|||||

Db 1596 TAAGTGGCTTTGAATTCGAACTCTACAGTATGCAGGATGACTATTACATATATTGGTATC 1655
Qy 2154 TCTCTGAATTCCTTTAGCATGGTTGATGTCAACATTTAGTGTGCGGATGGCTCTCAAA 2213
Db 1656 TCTCTGAATTCCTTTAGCATGGTTGATGTCAACATTTAGTGTGCGGATGGCTCTCAAA 1715
Qy 2214 TGGCAGAGGAAAGGATTAATGGAAGCAGCAGAAAGCCGCTAGTAGTAAAAAAACAAGA 2273
Db 1716 TGGCAGAGGAAAGGATTAATGGAAGCAGCAGAAAGCCGCTAGTAGTAAAAAAACAAGA 1775
Qy 2274 AAAAAAGAAAGTTTCGCCCATTTAGCCGAGAGATCACAATGAGCCAAAGCATATCAGAACA 2333
Db 1776 AAAAAAGAAAGTTTCGCCCATTTAGCCGAGAGATCACAATGAGCCAAAGCATATCAGAACA 1835
Qy 2334 TGTGTCTGGAATGTTTAAACCATGTTAGCATTTTGACATGGACGGCAAGTAGCTAAAC 2393
Db 1836 TGTGTCTGGAATGTTTAAACCATGTTAGCATTTTGACATGGACGGCAAGTAGCTAAAC 1895
Qy 2394 CGAAGTTTGTAGCTTGTAGTGAACAAGTTCGGTATGAACACAGGTTTGTCTCAATTCACA 2453
Db 1896 CGAAGTTTGTAGCTTGTAGTGAACAAGTTCGGTATGAACACAGGTTTGTCTCAATTCACA 1955
Qy 2454 GTGTGATGAACCCGCCGCCAGTGCATCTTACAGTTTCAAGGAAATGTCTGACCTCAATA 2513
Db 1956 GTGTGATGAACCCGCCGCCAGTGCATCTTACAGTTTCAAGGAAATGTCTGACCTCAATA 2015
Qy 2514 AATATAGCCCTCTCCCTCAGTCTCTGAACTGTATGTGGCAGCTAGTAAGCAGCTTCAAC 2573
Db 2016 AATATAGCCCTCTCCCTCAGTCTCTGAACTGTATGTGGCAGCTAGTAAGCAGCTTCAAC 2075
Qy 2574 AGGCAGAAATGATATTGGAATAATCTTAACCCGACCATGAGGTTTAATAGAATTTTAA 2633
Db 2076 AGGCAGAAATGATATTGGAATAATCTTAACCCGACCATGAGGTTTAATAGAATTTTAA 2135
Qy 2634 AGGTTGCCAAACCCAACTTTTGTGTTATGAAGTTTATGGCAGGAGGACACAAAAGGAAT 2693
Db 2136 AGGTTGCCAAACCCAACTTTTGTGTTATGAAGTTTATGGCAGGAGGACACAAAAGGAAT 2195
Qy 2694 CTAAGTTCCTCCCTGAAATTTGATTTCTCTCATTAATATTTTCCCTGTTGTAACCTG 2753
Db 2196 CTAAGTTCCTCCCTGAAATTTGATTTCTCTCATTAATATTTTCCCTGTTGTAACCTG 2255
Qy 2754 TTTTCAAGAGACTGGGAGGTGGCCATAAAGGGGCGAGAGTCTTCTTTCAAGACCCCACTCT 2813
Db 2256 TTTTCAAGAGACTGGGAGGTGGCCATAAAGGGGCGAGAGTCTTCTTTCAAGACCCCACTCT 2315
Qy 2814 TAGAGGCACATCAACAGGCTCCACATCAAGGAAAGTGAGATGTTCTTTGGGTAAACA 2873
Db 2316 TAGAGGCACATCAACAGGCTCCACATCAAGGAAAGTGAGATGTTCTTTGGGTAAACA 2375
Qy 2874 CTCATTATAGGATATCTTTAGTTTGCAGCCCTTATATGACATGAATGAAACTGCTGT 2933
Db 2376 CTCATTATAGGATATCTTTAGTTTGCAGCCCTTATATGACATGAATGAAACTGCTGT 2435
Qy 2934 TTTTAAAGTGGTTTATTTATGTTCCATGGAAGAAACTGGTCTTATTTGAAT 2981
Db 2436 TTTTAAAGTGGTTTATTTATGTTCCATGTAAGACACTGGGTTCCATTAT 2483

RESULT 3

AY102701

LOCUS

DEFINITION Mus musculus embryonic growth-associated protein EGAP mRNA, complete cds.

ACCESSION

AY102701

VERSION

AY102701.1

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

Wenzlau, J.M. and Weiser-Evans, M.C.M.

Db 538 ATTCCAGGGAAGAGTCAACAAGCTGCTGCTTTTGAGAGGAAGATTTCCAGTCAATG 597
Qy 1090 ACTTATGGATTAAATGGCTTAACAGTGTGACAGATCTTCGAGTTACAGGCATGCTAAAA 1149
Db 598 ACATATGGATTAAATGGCCACAGGTGTGACAGATCTTCGAGTTACAGGATGCTCAAA 657
Qy 1150 GATGTGAGGATGATGCAATGCAAGAGAGTAAGAGTACTCGAAGTCAACAAGGAAGAA 1209
Db 658 GATGTGAGGACGATGCAAGAGAGAGTAAGAGTACTCGAAGTCAACAAGGAAGAG 717
Qy 1210 AGAGATCCAGAAGTTGAATAGAACACCAACAAATGTTTAGCAGTATTTCAGCAGAGTGAAA 1269
Db 718 AGAGACCCAGAGTTGAATAGAACACCAAGAGTGTGCGCAGCGTTTCAGCAGAGTGAAG 777
Qy 1270 TTACTCGTGTGTTACTGACAGTGTCTATAGCCTTTTACTAAGAAAGACAGAGTGTGTT 1329
Db 778 TTACCCGAGTACTGCTGACAGTGTCTATAGCCTTTCACTAAGAAAGAGACAGTGTGTT 837
Qy 1330 GCAGAAGCTCAAAATTTGATGTTCAAGCAGCAGATCTTCTTCTGCCATTCATTAATCA 1389
Db 838 GCAGAGCTCAGAACTTGATGGTCCAGGCGCAGACCTTCTTCTGCCNTTCCACCTCA 897
Qy 1390 TTGCATCATGSCATCCAGGCTCAGATGSCACTACCAAGAGACATCCAAATATGATG 1449
Db 898 TTGCACCAAGGCTCAGGCTCAGATGSCACTACCAAGAGACATCCAAATATGATG 957
Qy 1450 GGTTTTGAACCCCTTGTGAACCGAGGCTACTCCACTACCTTCCCTCGATATGCAAAA 1509
Db 958 GGTTTTGAACCCCTTGTGAACCGAGATTAATCCACCCACTTCCCTCGTATGCAAAA 1017
Qy 1510 ATAAATAAAGGAAGAAATGCTGAACATTTTGCAGATTAATAGATAGATAAAACT 1569
Db 1018 ATAAATAAAGGAAGAAATGCTGAACATTTTGCAGATTAATAGATAGATAAAACT 1077
Qy 1570 GTCTGTGAGGTGTGAAATTTAAACAAATTTTCAATGTATCTGGAATTTTCTGGAATTT 1629
Db 1078 GTCTGTGAGGTGTGAAATTTAAACAAATTTTCAATGTATCTGGAATTTTCTGGAATTT 1137
Qy 1630 AGTGAAAGCTCAGATGTGTTCTTTCAAGATCTCTGTTAACAACCACTTTCTGAGTGGAT 1689
Db 1138 AGTGAAAGCTCAGATGTGTTCTTTCAAGATCTCTGTTAACAACCACTTTCTGAGTGGAT 1197
Qy 1690 AAGAAAAGCTTTTGAACTCATCTCATCAGACATGCTGAAGATGACACTTCGCTCT 1749
Db 1198 AAGAAAAGCTTTTGAACTCATCTCATCAGACATGCTGAAGATGACACTTCGCTCT 1257
Qy 1750 TTTGTGAGATCTTCCCGAGTCTTTCCCGAAGTCTACTATATATAATCAACAGGCTA 1809
Db 1258 TTTGTGAGATCTTCCCGAGTCTTTCCCGAAGTCTACTATATATAATCAACAGGCTA 1315
Qy 1810 AGAGCTGATCGACTCTTTGTTACTCATCTGTTGTGCGCCATCTCTGAGTCTTATTGAGA 1869
Db 1316 AGAGCTGATCGACTCTTTGTTACTCATCTGTTGTGCGCCATCTTTGAGTCTTATTGAGA 1375
Qy 1870 TCCATGAGATTAACAGGCTCGACAGAGATAGCTTGTGCTATATCTTGAAGATTTG 1929
Db 1376 TCCATGAGATTAACAGGCTCGGACAGAGACAACTTGTGCTACATCTTTGAAGAGTTG 1435
Qy 1930 CCACCTTGCAGGATGAGTTTATGACATTTTATTTAATAGGACAGAGAGTTGATGCGAGC 1989
Db 1436 CCACCTTGCAGGATGAGTTTATGACATTTTATTTAATAGGACAGAGAGTTGATGCGAGC 1472
Qy 1990 GCTTCACACCATGCTGTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2049
Db 1473 GCTTCACACCATGCTGTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1532
Qy 2050 CTGGGTCTTTTACATAACCTTCGATTTATGATACAGTACTCTTCTAGTGGCTTTGAAAT 2109
Db 1533 TTGGGTCTTACACAGCTCGGAGATTGATCCAGTACTGCTCAGCGGCTTTGATCT 1592
Qy 2110 GGAACCTACAGTATGACAGGATGACTATTAATATATATGATGATCTCTCTGAAATCTCTTTA 2169

Db 1593 GGAGCTCTACAGCATGCAAGAGTACTACTACATCTACTGGTACTCTCTCTGAGTTCTTGTA 1652
Qy 2170 CGCATGTTGATGTCACATTTGATTCGTGCGGATGGCTCTCAATGGCAGAGGAAGGAT 2229
Db 1653 CGCATGTTGATGTCACATTTGATTCGTGCGGATGGCTCTCAATGGCAGAGGAGAT 1712
Qy 2230 AATGGAAGCAGCAG 2289
Db 1713 AATGGAAGCAGCAG 1772
Qy 2290 CCATTTGAGCCGAGAGATCAATGAGCCAGAGATCAAGAGAGAGAGAGAGAGAGAGAG 2349
Db 1773 TCCATTTGAGCCGAGAGATCAATGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1832
Qy 2350 TAAAAACCATGTTGACATGTTGACATGTTGACATGTTGACATGTTGACATGTTGACATG 2409
Db 1833 CAAAACGATGTTGACATGTTGACATGTTGACATGTTGACATGTTGACATGTTGACATG 1892
Qy 2410 TAGTGAACAGTTGCGTATGAAACACAGGTTTGTCTCCATTTCAACAGTGTGATGACCCGCC 2469
Db 1893 CAGTGAGCAGGTCGGTATGAGCAGAGTTTGTCTCCGTTCAACAGCAGTGTGATGACGCCAC 1952
Qy 2470 GCCAGTGCATCTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAGCCCTCTCC 2529
Db 1953 ACCAGTGCATCTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAGCCCTCTCC 2012
Qy 2530 TCAGTCTCTGAACTGTATGTTGTCAGCTAGTAAAGCACTTTCAACAGGCAAAATATGATTT 2589
Db 2013 TCAGCTCCAGAGCTGTATGTTGTCAGCTAGTAAAGCACTTTCAACAGGCAAAATATGATTT 2072
Qy 2590 GGAATAATTCCTAACCCGACCATGAGGTTAATAGAAATTTTAAAGTTTCCCAAAACCCAA 2649
Db 2073 TGAGAGCATCCCCAATGCAGACCCGGAGGTCAGCAGGATCTTAAAGGTTGCCAAGCCTAA 2132
Qy 2650 CTTTGTGTTTATCAAGTTTATGTCAGGAGACACAAAGAGGATCTTAAAGTTTCCCTCTGA 2709
Db 2133 CTTTGTGTTTATGAGCTCTTACAGGAGACACAAAGAGGATCTTAAAGTTTCCCTCTGA 2192
Qy 2710 ATTGATTTCTCTGCTCATAAATATTTTCTGTTGTGAAACTTTGTTGAGAGAGACTGGG 2769
Db 2193 ATTGATTTCTCTGTTTCAAAATATTTTCTGTTGTGAAACTTTGTTGAGAGAGACTGCA 2252
Qy 2770 GAGTGGCCATAAAGGGCAGAGTCTTCTTTCAGACCCAACTCTTAGAGGGCAGATCAC 2829
Db 2253 AAGATGACCAT-----GTCTACTCTGGGGTACACCA 2283
Qy 2830 AGCTCCACATCACGGGAGTGAATGTTCTGGGTAAACAACTCATTTATAAGGA--- 2886
Db 2284 GTGTGCCACACACGGTGAATAAGTGGCTCTCTTGAATGACAGCTTGTTCAGAGAGAC 2343
Qy 2887 -ATATTTAGTTGACAGCTTTATATGACATGAATGAATACTGCTGTTTTTAAAGTGGTT 2945
Db 2344 TTCAGTATCAGTACAAACCCCTTACCTGGAGAGAACG-AAACGCTGTTTTTAAAGTGGTT 2402
Qy 2946 TATTATGTTCCATGGAAGAACTGGTCTTATTTGAATGCATGATGAACGTTATATGGTTT 3005
Db 2403 TGTATATCTTCCATGGGTGACGGG-----CTGGAATGCATGGTGAACGTTATATGGTTT 2458
Qy 3006 TATTACAGATTTAATCACAATCATTTTATGAATGATTGATGAAATAGTGTTTATA 3065
Db 2459 TATTACAGATTTAATGTAATACTTTTAAATGATGATTTGTG-AAATAGTGTTTGTA 2517
Qy 3066 AGGTTAATAAATTTCTTGACAAAA 3091
Db 2518 AAGGTTAATAAATTTCTTGACAAAA 2543

RESULT 6
AK025266
LOCUS
DEFINITION
ACCESSION
VERSION

AK025266 1839 bp mRNA linear PRI 13-SEP-2003
Homo sapiens cDNA: FLJ21613 fis, clone COL07381.
AK025266
AK025266.1 GI:10437744

KEYWORDS	oligo capping; fis (full insert sequence).	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	Kawabata, A., Hiki, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okizaki, R., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.	
TITLE	NEDO human cDNA sequencing project	
REFERENCE	Unpublished	
AUTHORS	2 (bases 1 to 1839)	
TITLE	Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T., Shibahara, T., Tanaka, T. and Nakamura, Y.	
JOURNAL	Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail: flicdn@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5286, Fax: 81-3-5449-5416)	
COMMENT	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- and 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).	
FEATURES	Location/Qualifiers	
source	1..1839	
	/organism="Homo sapiens"	
	/mol_type="mRNA"	
	/db_xref="taxon:9606"	
	/clone="COL07381"	
	/tissue_type="colon"	
	/clone_lib="COL"	
	/note="Cloning vector pME18SFL3"	
	134..1714	
	/note="unnamed protein product"	
	/codon_start=1	
	/protein_id="BAB15097.1"	
	/db_xref="GI:10437745"	
	/translation="MVKASVDDDDSGWELSMPEKMSNTNWVDITDPEACRELKLGELHDKLFGLEFMSAIEMWDPKMDAGMIGNVRKLVNFEQAIKDGITIKIDLTAPELIGMDTCCLITWLEHSLAQVFTCLYINPDPIEDPAMKAPALGILKIDIA REKNKAAYPEEPQSMYTGFKMANSVTDLRVTGMLKDVEDDQRRVKSTRSQGEE RDPVELEHQCLAVFSRKVETRVLLITVLAFTKETSABVAQKLMQARADLLSALH NSLHVGQAQNDTTKGDPHPIMWFEPLVNQRLLPPTFPFYAKIKREEMVNYFARLID RIKTVCEVNI.TNLHCLIDFFCEFSQSPVLSRLSLQTLTFVLVDNKKYFGTHLMQDMV KDALRSFSPVLPKCYLVNHOAKCIDSFVTHVRPFCSLIQIHGHNRAORDKL GHILEEFATLQDEAEKVDAALHTMLLKQEPQRLAWLGTWLVYHNLRIIMIQYLLSGF ELELYSMHEILLHIVSL"	
CDS		
	/protein_id="BAB15097.1"	
	/db_xref="GI:10437745"	
	/translation="MVKASVDDDDSGWELSMPEKMSNTNWVDITDPEACRELKLGELHDKLFGLEFMSAIEMWDPKMDAGMIGNVRKLVNFEQAIKDGITIKIDLTAPELIGMDTCCLITWLEHSLAQVFTCLYINPDPIEDPAMKAPALGILKIDIA REKNKAAYPEEPQSMYTGFKMANSVTDLRVTGMLKDVEDDQRRVKSTRSQGEE RDPVELEHQCLAVFSRKVETRVLLITVLAFTKETSABVAQKLMQARADLLSALH NSLHVGQAQNDTTKGDPHPIMWFEPLVNQRLLPPTFPFYAKIKREEMVNYFARLID RIKTVCEVNI.TNLHCLIDFFCEFSQSPVLSRLSLQTLTFVLVDNKKYFGTHLMQDMV KDALRSFSPVLPKCYLVNHOAKCIDSFVTHVRPFCSLIQIHGHNRAORDKL GHILEEFATLQDEAEKVDAALHTMLLKQEPQRLAWLGTWLVYHNLRIIMIQYLLSGF ELELYSMHEILLHIVSL"	
ORIGIN		
Query Match	57.1%; Score 1766.6; DB 9; Length 1839;	
Best Local Similarity	98.1%; Pred. No. 0;	
Matches 1828; Conservative	0; Mismatches 9; Indels 26; Gaps 3;	
QY	423 ATACGCATCGTGCACGCTGCGGCTCGGCTCGGCTGAGAGGGAGGGGGCGGCGGC 482	663 TGAAGAAGCTTGTGAGAAATTAAAGTTGGGAGAACTACTTTCATGATAGCTATTGGTCT 722
DB	1 ATACGCATCGTGCACGCTGCGGCTCGGCTGAGAGGGAGGGGGCGGCGGC 60	241 TGAAGAAGCTTGTGAGAAATTAAAGTTGGGAGAACTACTTTCATGATAGCTATTGGTCT 300
QY	483 CGAGCGCGCTGTTATTTTCGTGTCGCGAGTGCCTGCGGCGGCGGCTGACCA 542	723 TTTTGAAGCCATGCTCTCTATTGAAATGATGGATCCCAAGATGATGCTGGCATGATGG 782
DB	61 CGAGCGCGCTGTTATTTTCGTGTCGCGAGTGCCTGCGGCGGCGGCTGACCA 120	301 TTTTGAAGCCATGCTCTCTATTGAAATGATGGATCCCAAGATGATGCTGGCATGATGG 360
QY	543 AGAAGTAGCATTAATGTTATGAAAGCTTCTAGATGATGACGATTCAGGATGGAGCT 602	783 AAACCAAGTTAATCGAAAGTTCTCAATTTTGAACAAGCTATCAAGATGCGCATTTAA 842
DB	121 AGAAGTAGCATTAATGTTATGAAAGCTTCTAGATGATGACGATTCAGGATGGAGCT 180	361 AAACCAAGTTAATCGAAAGTTCTCAATTTTGAACAAGCTATCAAGATGCGCATTTAA 420
QY	603 CAGTATGCGAGAAATGCGAAAGCAATACAACTGGGTGGACATTTACCAAGATT 662	843 AATTAAAGATCTCACTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 902
DB	181 CAGTATGCGAGAAATGCGAAAGCAATACAACTGGGTGGACATTTACCAAGATT 240	421 AATTAAAGATCTCACTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
		903 GATAACGTGGTTAGAAAGCCATTCACTGGGCAAGACAGTATTTACGTGCTTTCATTTCA 962
		481 GATAACGTGGTTAGAAAGCCATTCACTGGGCAAGACAGTATTTACGTGCTTTCATTTCA 540
		963 TAATCCAGACTTTATAGAGATCTGCTATGAAAGCTTTTCTCTGCTGCTGCTGCTGCTGCT 1022
		541 TAATCCAGACTTTATAGAGATCTGCTATGAAAGCTTTTCTCTGCTGCTGCTGCTGCTGCT 600
		1023 CTGTGACATTCGAAAGGAAAAAGTAAATAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1082
		601 CTGTGACATTCGAAAGGAAAAAGTAAATAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
		1083 GTCATGACCTTATGAAATTTAAATGGCTTAAACAGTGTGACAGATCTTTCAGGTTACAGGAT 1142
		661 GTCATGACCTTATGAAATTTAAATGGCTTAAACAGTGTGACAGATCTTTCAGGTTACAGGAT 720
		1143 GCTAAAGATGCTGGAGGATGACATGCAAGAGAGTAAAGAGTACTCCAAAGTCCGCAAGG 1202
		721 GCTAAAGATGCTGGAGGATGACATGCAAGAGAGTAAAGAGTACTCCAAAGTCCGCAAGG 780
		1203 AGAAGAAGAGATCCAGAAAGTTGAACTAGAACACCAATGTTTAGCAGTATTTCAGCAG 1262
		781 AGAAGAAGAGATCCAGAAAGTTGAACTAGAACACCAATGTTTAGCAGTATTTCAGCAG 840
		1263 AGTGAATTTACTCGTGTGTTACTGACAGTGTGTTATAGCCTTTTACTTAAGAAAGACAC 1322
		841 AGTGAATTTACTCGTGTGTTACTGACAGTGTGTTATAGCCTTTTACTTAAGAAAGACAC 900
		1323 TGCTGTTGCAAGCTCAAAATTTGATGTTTCAAGCAGCAGATCTTCTTTCTGCGCATTTCA 1382
		901 TGCTGTTGCAAGCTCAAAATTTGATGTTTCAAGCAGCAGATCTTCTTTCTGCGCATTTCA 960
		1383 TAAATTCATTTGATGATGCGATCCAGGCCCCAGAAATGATCTACAAAGAGAGATTCACCAAT 1442
		961 TAAATTCATTTGATGATGCGATCCAGGCCCCAGAAATGATCTACAAAGAGAGATTCACCAAT 1020
		1443 TATGATGGGTTTGAACCCCTTGTGAACAGAGGCTACTTCCACCTACTCTTCCCTCGATA 1502
		1021 TATGATGGGTTTGAACCCCTTGTGAACAGAGGCTACTTCCACCTACTCTTCCCTCGATA 1080
		1503 TGCAAAAATAATTAAGAGGAGAAATGCTGAACTATTGTTTGAAGATTAATAGATAGAT 1562
		1081 TGCAAAAATAATTAAGAGGAGAAATGCTGAACTATTGTTTGAAGATTAATAGATAGAT 1140
		1563 AAAAATCTGCTGTGAGGTTGTGAATTTAAACAAATTTACATGATGATCTGCTGGAATTTTCTG 1622
		1141 AAAAATCTGCTGTGAGGTTGTGAATTTAAACAAATTTACATGATGATCTGCTGGAATTTTCTG 1200
		1623 TGAATTTAGTGAACAGTCCACCATGTTCTTTTCAGATCTCTGTTTACAAACCATTTCTCT 1682
		1201 TGAATTTAGTGAACAGTCCACCATGTTCTTTTCAGATCTCTGTTTACAAACCATTTCTCT 1260
		1683 GGTGATAACAAAAGGCTTTTGGAACTCATCTCATGCAAGACATGTTGGAAGATGCACT 1742
		1261 GGTGATAACAAAAGGCTTTTGGAACTCATCTCATGCAAGACATGTTGGAAGATGCACT 1320
		1743 TCGGTCTTTTGTGATGCTCCGAGTGTCTTCCCAAGTGTCTCTATATATATATATAC 1802

```
Db 1321 TCGGCTTTGTCAG--TCTCCGGTGTCTTCCGCCAGTGTACTATATATATATAC 1378
QY 1803 CAGGCTAAGGACTGTATCGACTCTTTGTTACTACATGTGTTCCGGCAATCTGTAGTCTT 1862
Db 1379 CAGGCTAAGGACTGTATCGACTCTTTGTTACTACATGTGTTCCGGCAATCTGTAGTCTT 1438
QY 1863 ATTACATCCATGACATACAGGCTCGACAGAGATAGCTTGGTCATATCTTGTAG 1922
Db 1439 ATTACATCCATGACATACAGGCTCGACAGAGATAGCTTGGTCATATCTTGTAG 1498
QY 1923 GAATTTGCCACCTTGCAGATGAGTTATGACATTTTATTTAATAGGACAGAGAAGTGTG 1982
Db 1499 GAATTTGCCACCTTGCAGATG-----AGGACAGAGAAGTGTG 1535
QY 1983 ATGACGCGCTTACACCATGCTTTGAACAGAGAACCCCAAGGCAACATTTGGCTGT 2042
Db 1536 ATGACGCGCTTACACCATGCTTTGAACAGAGAACCCCAAGGCAACATTTGGCTGT 1595
QY 2043 TAGGTACCTGGTCCCTTTACCAATACCTTGCATATGATACAGTACCTTCAAGTGGCT 2102
Db 1596 TAGGTACCTGGTCCCTTTACCAATACCTTGCATATGATACAGTACCTTCAAGTGGCT 1655
QY 2103 TTGAATTTGAACCTTACAGTATGACAGAG--TACTATTACATATATTTGGTATCTCTCTGAA 2161
Db 1656 TTGAATTTGAACCTTACAGTATGACAGAGACTATTACATATATTTGGTATCTCTCTGAA 1715
QY 2162 TTCCTTTAGCATGTTGATGATCAACATGAGTCGTCGCGATGCTCTCAATGGCAGAG 2221
Db 1716 TTCCTTTAGCATGTTGATGATCAACATGAGTCGTCGCGATGCTCTCAATGGCAGAG 1775
QY 2222 GAAAGATATATGAAGCAGCAGAGAGGCGGTAGTAGTAAATTAACAAAGAAAAAAG 2281
Db 1776 GAAAGATATATGAAGCAGCAGAGAGGCGGTAGTAGTAAATTAACAAAGAAAAAAG 1835
QY 2282 AAA 2284
Db 1836 AAA 1838

RESULT 7
BC053286 2467 bp mRNA linear VRT 07-OCT-2003
LOCUS Danio rerio cDNA clone MGC:64157 IMAGE:6797204, complete cds.
BC053286
ACCESSION BC053286.1 GI:31418975
VERSION MGC.
KEYWORDS Danio rerio (zebrafish)
SOURCE Danio rerio
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 2467)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L.D., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.P., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Dietlenon,M., Soares,M.B., Donald,M.F., Casavant,T.L.,
Schetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullaby,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Vialalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smallos,D.E.,
Schnerker,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
```

```
JOURNAL MEDLINE
PUBMED 12477932
REFERENCE 2 (bases 1 to 2467)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgabs.r@mail.nih.gov
Tissue Procurement: Leonard I. Zon, M.D.
cDNA Library Preparation: Invitrogen Corp
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 117 Row: n Column: 11
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, similarity but not identity to protein.
FEATURES
Location/Qualifiers
1..2467
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="MGC:64157 IMAGE:6797204"
/tissue type="kidney, zebrafish"
/clone_lib="NCI CGAP_ZKId"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6.1"
58..2232
/codon_start=1
/product="Unknown (protein for MGC:64157)"
/protein_id="AAH53286.1"
/db_xref="GI:31418976"
/translation="MVKSSVEEEEGNGLGIPEKRNANNWVDTVEFKGACKELKL
GELHDKUFLGFEAMSAIEMDPKMDAGINQVARKVLFDPQAVKDEAIRVKDLSIP
ELIGIMDTFCCLITWLEHSLAQVFTCLYVHNPDLIQDPALFALGILKICDIAR
EKVKAAPVEEDFOATYGFKMANVDTLRTVGLKQVDELOKRVKSTRQSGOR
DPEVLDHQOCLALFSRVKFTLLLSALISFTKKTSAVSEAKLMSQAADLLPAVHA
TTQYGIQSONDTTGDHPIMMGFEPLVNO RLPPFPFYAKILIKREWNYSKILIER
IKSVCEVINITNLHSLIDFFCFSESPCVLSRLQTTFLIDNKKVFGTHLMQDMIK
DALRCFVSPVLSSKCSLNHNOAKDYIDSFVTHCTPFCSLIQHGNRAQRDKLIG
HLEEFATLQDEAEKVALHGLMLKLEPQROHLACLGTWILYHLNLRIMQVLLSGFE
LLEYSMHYYIYVYLSFPLVAMLSLTSRADSSQMAERILEOLKVRSSKSKKKK
KARPLSKTMSQAYONCAGMYKTMIALDMDKVKPKQFELDSQVRYEHRFAPNS
VTPFPVHYIQKEMSCLKYKNPPRSLADLYMAASKHQQAOKLLENVTSPDAENVRI
LKVAPKNIVMKLLAGGHKKTKALPEFDFAHKKYFPVVKIL"
139..693
/note="Mak10; Region: Mak10 subunit, NatC
N(alpha)-terminal acetyltransferase. NatC
N(alpha)-terminal acetyltransferases contains Mak10p,
Mak3p and Mak3p subunits. All three subunits are
associated with each other to form the active complex"
/db_xref="CDD:pfam04112"
misc_feature
Query Match 40.9%; Score 1264.8; DB 5; Length 2467;
Best Local Similarity 74.3%; Pred. No. 5.9e-276;
Matches 1648; Conservative 0; Mismatches 542; Indels 28; Gaps 3;
QY 543 AGAAGTAGGCATAATGGTATGAAAGCTTCTGTAGATGATGACGATTCAGGATGGAGCT 602
Db 45 AANAGCTGTCTCCATGGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 104
```


Qy	603	CAGTATGCCAGAAAATGGAGAAAGCAATACAAACTGGGTGGACATTACCCAGATTT	662	Qy	1683	GGTGGATAACAAAAGGTCTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACT	1742
Db	105	TGGGATCCAGAGAGAT--GAGAAACAATGGCCAACTGGGTGGACGTACCCAGGATTT	161	Db	1182	GATCGACAACAAGAGGTGTTTGGACTCATCTGATGAGGATATGATTAAGATGCGCT	1241
Qy	663	TGAAGAAGCTTGTCAGAAATTAAGTTGGGAGAACTACTCATGATAAGCTATTGGTCT	722	Qy	1743	TCGGTCTTTTGTTCAGATCCTCCGAGTGTCTTCCGCCAAGTGTACCTATATAAATACAC	1802
Db	162	TAAAGGAGCTGTAAAGNACTCAAACTTGGCGAGTGTCTCATGACAAGCTATTTCGTCT	221	Db	1242	CGCGTGTTTTGTTCAG--TCCACCTGTTCTTTCTCTCAAAATGCAAGTGTGAATAAATACAC	1299
Qy	723	TTTTGAAGCCATGTCGTCTATGTAATGATGGATCCCAAGATGGATGTCGATGATGG	782	Qy	1803	CAGGCTAAGGACTGTATTCGACTCTTTGTTACTCATCTGTGTTCGGCCATCTCTGTAGTCTT	1862
Db	222	GTTCGAAGCCATGTCGCGCAATGAGATGATGGATCCGAAGATGGATGCTGGAATGATCG	281	Db	1300	CAGGCTAAGGATTTATTCGACTCTTTTGTTCACACACTGCACGCGCGCGTCTTCGACGCTC	1359
Qy	783	AAACCAAGTTAATCGAAAGTTCTCAATTTTGAACAGCTATCAAGGATGCACTATTA	842	Qy	1863	ATTTCAGATTCATGACATATAACAGGGCTCGACAGAGAGATAAGCTTGGTTCATATTCTTCAG	1922
Db	282	AAATCAAGTCAACGTAAGTCTCAACTTTTGAACAGCTTTAAGGATGAAGCCATACG	341	Db	1360	ATCCAGATTTCAGGACACACCGAGCCCGACAGAGAGACAACCTGGGTTCATATTCTGGAG	1419
Qy	843	AAATTAAGATCTCACTTTCCTGCAACTGATAGGATTTATGATACATGTTTTTCTGTTT	902	Qy	1923	GAATTTGCCACCTTTCAGGATGAGTTTATGACATTTTTTATTAATAGGCAGAGAAGTTG	1982
Db	342	GGTGAAGATCTAAGTATTCCTGAGCTCATCGGATCATGACACATGTTTCTGCTGCT	401	Db	1420	GAGTTGCCACACTTCAGGATG-----AGGCAGAGAAGGTGG	1456
Qy	903	GATTAAGCTGTAGAGCCATTCATCGGACAGACAGATATTAGTGCCTTTACATTC	962	Qy	1983	ATGCAGCGCTTCACACCATGCTGTTGAAAAGGAAACCCCAAGCAACATTTGGCCTGTT	2042
Db	402	GATCACGTGGTGGAGGGGCAATCTCTGGCTCAGACCGCTCTTACATGCTGTGACGTTCA	461	Db	1457	ACGCTGCTCTTCACGGTCTGCTGATGAAGCTGGAGCCGACGCGCAGCATTTAGCCTGTC	1516
Qy	963	TAATCCAGACTTTATAGAGATCCTGCTATGAAGCTTTTCTGCTGGGAATCTTGA	1022	Qy	2043	TAGGTACTCGGTCTTTTACATAACCTTCGATTTATGATACACTACCTCTTAAGTGGCT	2102
Db	462	TAATCCGATCTGATTAAGACCTTCGCTCAAGCCCTTCGCTCTGGGATCTCTGA	521	Db	1517	TGGGCACTGGATCTCTACCAACCTGCGCATCATGATCCAGTACTCTGCTCAGCGCT	1576
Qy	1023	CTGTGACATTCAGAGGAAAGTAAATAAAGCTGCTGTTTTTGAAGGAAGATTTTCA	1082	Qy	2103	TTGAATTTGGAACCTCAAGTATGCAAGTATGCACTATTAATACATATATGTTATCTCTGAAT	2162
Db	522	CTGTGACATCGCCGTGAAAGTCAACAAAGCTGCGGTGTTTGAAGGAAGATTTTCA	581	Db	1577	TTGAGCTGGAGCTCTACAGCATGCACGAATACTATTACATCTACTGGTATCTGTCAAGT	1636
Qy	1083	GTCAATGACTTATGATTTAAATGGCTAAACAGTGTGACAGATCTTCAGTTACAGGAT	1142	Qy	2163	TCCTTTACGATGTTGATGTCACATTTGATGTCGCGGTGCTCTCAATATGCGCAGG	2222
Db	582	GGCCATGACTTACCGCTTTAAATTTGGCCAAACAACTGATCTGATCTGAGAGTGA	641	Db	1637	TCCTTTATGCGTGGCTGATGTCGACGTTGAGTTCGAGCCGACAGCTCTCAGATGCGGAG	1696
Qy	1143	GCTAAAGATCTGGAGGATGACATGCAAGGAAGATTAAGAGTACTCGAAGTCGACAA	1202	Qy	2223	AAAGGATAATGGAAGACGACAGAAAGCGCGTGTAGTAAATAAAGGAAAAAAGAA	2282
Db	642	GCTGAAGATGTTGAGGATGAATTCAGAGGAAGTAAAGATACCCGATGACAGG	701	Db	1697	AGAAATCTTGGAGGACGCTCAAGTGGCGACGACGACGACGACGACGACGACGACG	1756
Qy	1203	AGAAAGAGATCCAGAGTTGAACCTAGAACACCAACAATGTTTTAGCAGTATTCAG	1262	Qy	2283	AAGTTCGCCCATTTAGCCGAGAGATCAATAGCCAAAGCATATCAAGAACATGTGTGCTG	2342
Db	702	TGAACGCGAGATCCTGAGTGGAACTGGATCATCAGCAGTGTGGCTGTTTCA	761	Db	1757	AAGCTGCTCTCTGAGTAAAGAGATCACCATGAGCCAGGCTTACCAAGAACATGTGTGCTG	1816
Qy	1263	AGTGAATTTACTGTTGTTACTGACAGTGTCTTATAGCTTTTACTAAGAAAGACG	1322	Qy	2343	GAATGTTTAAACCATGGTAGCATTTGACATGGAGCGGCAAGTACGTAACCGAAGTTG	2402
Db	762	GGTTAAATTTCACTGCTGCTGCTCAGCGCGCTAATCTCTTCACCAAAAGAGACG	821	Db	1817	GGATGTAAGACGATGATCGCGTGGATATGGATCGAAAGGTGCGCAAACTCAGTTTG	1876
Qy	1323	TGCTGTTGCAAGCTCAAAATTTGATGGTTCAAGCAGCAGATCTTCTTCTGCCATTCA	1382	Qy	2403	AGCTTGATAGTGAACAAAGTTCCGTTATGAACACAGGTTTGTCTTCCATTCAACAGTGTGATGA	2462
Db	822	TGCTGTGAGCGAAGCAGAGAGTGTGAGTCAAGCTGCTGATTTACTGCTTCAGTCCA	881	Db	1877	AGCTGACAGTGGACAGTGGCTACGAGCATCGCTTCGCCCTTTTCAACAGTGTGCTCA	1936
Qy	1383	TAAATTCATGTCATGATGTCATCGAGCCAGAAATGATATCAAAAAGGAGATCAATCA	1442	Qy	2463	CCCGCGCGCAGTGCATCTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAGCC	2522
Db	882	TGCTACCATACAGTACGGCATACAGTGCAGAACGACACTACCAAAAGGAGATCA	941	Db	1937	CACCACCACAGTTTCACTACATCCAGTTTAAAGGAAATGTCCGATTTGAAGAAGTACATC	1996
Qy	1443	TATGATGGTTTGAACCCCTGTGAACCCAGAGGCTACTTCCACTTCCCTCGATA	1502	Qy	2523	CTCCTCTCAGTCTCTGTAAGTGTATGTCGAGCTAGTAAAGCAGCTTTTCAACAGGCAAAA	2582
Db	942	CATGATGGTTTGAAGCTCTGGTTTAAACAGCGCTGCTCCACTAGTTTCCCGGATA	1001	Db	1997	CTCCTCTCGCTCAGCGGATCTCTATATGGTTCAGTAAACATTTCCAGCAGGCGCAAC	2056
Qy	1503	TGCAAAATAATTAAGGGAAGAAATGGTGAATTTTTCGAAGATTAATAGATAGAT	1562	Qy	2583	TGATTTTGGAAAATATTCTTAACCCGAGCATGAGGTTAATAGAAATTTTAAAGGTGCGCA	2642
Db	1002	TGCCAAGATCATCAAGCAGAGAGATGTCATCTCTCAGCAAACTCATTTAGCGCAT	1061	Db	2057	TCCTTACTGGAAAATGTCCACGAGCTTGTATGAGGTTTAAATCGTATCTCTTAAAGTGGCA	2116
Qy	1563	AAAACTGCTGTGAGGTTGAAATTAACAAATTTACATTTGATCTGAGATTTTCTG	1622	Qy	2643	AACCAACTTTGTTTATGAGTTTATGAGTTTATGCGAGGACACAAAAGGAAATCTAAAGTTC	2702
Db	1062	TAAGAGCGGTGTGTGAGGTCATCAACATCAACAACTGACAGCATCTCTGGATTTCTG	1121	Db	2117	AACCAATAATTTGTTTAAATGAGCTGCTGGAGGACACAAAGAGAGACCAAGGCC	2176
Qy	1623	TGAATTTAGTGAACAGTCACATGTTCTTTTCAAGATCTCTGTTTACAAACCATTTCT	1682	Qy	2703	CTCCTGAATTTGATTTCTCTGCTCATAAATATTTTCTGTTGTGAAACTTTGTTTGA	2760
Db	1122	TGAATTTAGTGAGAGTCTTCAATGTGTGCTGTAGATCTCTACTGACAGACAGTTTCT	1181	Db	2177	TTCTCTGAATTTGATTTCTCTGCTCACAATACTTTCCCATAGTCAAAATCTCTATGAGA	2234


```

RESULT 8
AK026296
LOCUS Homo sapiens cDNA: FLJ22643 fis, clone HSI07031. PRI 13-SEP-2003
DEFINITION
ACCESSION AK026296
VERSION AK026296.1 GI:10439113
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H.,
Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure Analysis, Human
Genome Center, Shirokane-dai, 4-6-1, Minato-Ku, Tokyo 108-8639,
Japan (E-mail:flcdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source
1. .997
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HSI07031"
/tissue_type="human small intestine"
/clone_lib="HSI"
/notes="cloning vector pME18SFL3"
15. .650
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB15435.1"
/db_xref="GI:10439113"
/translations="MHEYYIYVYLSEFLYAWLMSTLSRADGSGMAEERIMEBQOKGR
SSKTKKKKKVRLPSREITMSQAYQNMCGMKTWAFDMDGKVRKPKFLDSEQVRY
EHRFAPNSVMTPPVHYLQPKEMSLNKYPPQSPPELYVAASKHFQQAQKMLENIP
NPDEHVRILKVAKNFVNMKLLAGGHKESKVPPEFDFSAHKYFFPVVKLV"
CDS
Query Match 31.9%; Score 988; DB 9; Length 997;
Best Local Similarity 100.0%; Pred. No. 3 4e-213;
Matches 988; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2109 TGAATCTTACAGTATGCACAGTACTATTACATATATTGGTATCTCTCTGAATTCCTTT 2168
DB 1 TGAATCTTACAGTATGCACAGTACTATTACATATATTGGTATCTCTCTGAATTCCTTT 60
QY 2169 ACCGATGTTGATGTCAACATTGAGTCGTCGCCATGCTCTCAATGGCAGAGAAAGGA 2228
DB 61 ACGCATGTTGATGTCAACATTGAGTCGTCGCCATGCTCTCAATGGCAGAGAAAGGA 120
QY 2229 TAATGGAAGACGACGAGAAAGCGGTAGTAGTAAATAAACAAGAAAAAGAAAGTTC 2288
DB 121 TATGGAAGACGACGAGAAAGCGGTAGTAGTAAATAAACAAGAAAAAGAAAGTTC 180
QY 2289 GCCATTGACCGGAGAGATCACAAATGAGCCAGCATATCAGACATGTGTGCTGGAATGT 2348
DB 181 GCCATTGACCGGAGAGATCACAAATGAGCCAGCATATCAGACATGTGTGCTGGAATGT 240

```

```

QY 2349 TTAATAACCATGTTAGCATTTGACATGACGGCAAGTAGCTAAACCGAAGTTTGAGCTTG 2408
DB 241 TTAATAACCATGTTAGCATTTGACATGACGGCAAGTAGCTAAACCGAAGTTTGAGCTTG 300
QY 2409 ATAGTGAACAAGTTCGGTATGAACACAGGTTTCTCATTCAACAGTGTGATGACCCCGC 2468
DB 301 ATAGTGAACAAGTTCGGTATGAACACAGGTTTCTCATTCAACAGTGTGATGACCCCGC 360
QY 2469 CGCAGTGCACACTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAGCCCTCTC 2528
DB 361 CGCAGTGCACACTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAGCCCTCTC 420
QY 2529 CTCAGTCTCTGACACTGTATGTGGCAGCTAGTAGACACATTTCAACAGGCGAATAATGATAT 2588
DB 421 CTCAGTCTCTGACACTGTATGTGGCAGCTAGTAGACACATTTCAACAGGCGAATAATGATAT 480
QY 2589 TGGAAAAATATTCTCTAAACCGGACCATGAGGTAAATAGAAATTTTAAAGGTTGCCAAACCCA 2648
DB 481 TGGAAAAATATTCTCTAAACCGGACCATGAGGTAAATAGAAATTTTAAAGGTTGCCAAACCCA 540
QY 2649 ACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGTTCTCTCTG 2708
DB 541 ACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGTTCTCTCTG 600
QY 2709 AATTTGATTTCTCTGCTCATATAATATTTTCTGTTGTGAACTTTGTTGAGAGACTGG 2768
DB 601 AATTTGATTTCTCTGCTCATATAATATTTTCTGTTGTGAACTTTGTTGAGAGACTGG 660
QY 2769 GGAGGTGGCCATAAAGGGGACAGTCTCTTTTTCAGACCCCACTCTTTAGAGGGACATATCAC 2828
DB 661 GGAGGTGGCCATAAAGGGGACAGTCTCTTTTTCAGACCCCACTCTTTAGAGGGACATATCAC 720
QY 2829 CAGGCTCCACATCACCGGAAAGTGGATGGAATTTCTGGGTAAACAATCATATATAAGGAAT 2888
DB 721 CAGGCTCCACATCACCGGAAAGTGGATGGAATTTCTGGGTAAACAATCATATATAAGGAAT 780
QY 2889 ACTTTTAGTTTGCACAGCTTATATGACATGAATGAACCTGCTGTTTAAAGTGGTTTAT 2948
DB 781 ACTTTTAGTTTGCACAGCTTATATGACATGAATGAACCTGCTGTTTAAAGTGGTTTAT 840
QY 2949 TATGTTCCATGGAAGAACTGGTCTTATTGAATGCAATGATGAACCTGTTATATGCTTTTAT 3008
DB 841 TATGTTCCATGGAAGAACTGGTCTTATTGAATGCAATGATGAACCTGTTATATGCTTTTAT 900
QY 3009 TACAGATTAAATCACAATAATCAATTTTATGAATGATGAGTGAATAATAGTGTATATAAG 3068
DB 901 TACAGATTAAATCACAATAATCAATTTTATGAATGATGAGTGAATAATAGTGTATATAAG 960
QY 3069 GTTAATAAATTTCTTGCACAAAAA 3096
DB 961 GTTAATAAATTTCTTGCACAAAAA 988

```

```

RESULT 9
AK056059
LOCUS Homo sapiens cDNA FLJ31497 fis, clone NT2NE2005317, highly similar
DEFINITION to rattus norvegicus corneal wound healing related protein mRNA.
ACCESSION AK056059
VERSION AK056059.1 GI:16551204
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T.,
Kodaira,H., Furuya,T., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K.,
Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1

```

Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,
 Negahari, K., Masuo, Y., Nagai, K. and Isogai, T.
 NEDO human cDNA sequencing project
 Unpublished
 2 (bases 1 to 1796)
 Isogai, T., Otsuki, T. and Sugiyama, T.
 Direct Submission
 Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
 Genomics Laboratory; 1332-3 Yana, Kisarazu, Chiba 292-0812, Japan
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:
 Research Association for Biotechnology (RAB); cDNA library
 construction: Helix Research Institute (HRI) (supported by Japan
 Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
 HRI, and Biotechnology Center, National Institute of Technology and
 Evaluation; clone selection for full insert sequencing: RAB and
 HRI.

FEATURES

Location/Qualifiers
 1..1796
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="NT2NE2005317"
 /cell_line="NT2"
 /cell_type="teratocarcinoma"
 /clone_lib="NT2NE2"
 /note="cloning vector: pME18SFL3-mRNA from NT2 neuron
 after the differentiation of NT2 neuronal precursor
 cells."

ORIGIN

Query Match 28.6%; Score 885.4; DB 9; Length 1796;
 Best Local Similarity 99.9%; Pred. No. 7.2e-190;
 Matches 886; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 549 AGGCATATGCTTATGAAGCTTCTGTAGATGATGACCATTCAGGATCGGAGCTCAGTAT 608
 Db
 345 AGGCATATGCTTATGAAGCTTCTGTAGATGATGACCATTCAGGATCGGAGCTCAGTAT 404
 Qy
 609 GCCAGAAAAATGGAGAAAAACAATACAACTGGGTGGACATTTACCCAAAGATTTTGAAGA 668
 Db
 405 GCCAGAAAAATGGAGAAAAACAATACAACTGGGTGGACATTTACCCAAAGATTTTGAAGA 464
 Qy
 669 AGCTTGTGCGAAATTAAGTTGGGAGAACTCTCATCATAGCTATTTGGTCTTTTGA 728
 Db
 465 AGCTTGTGCGAAATTAAGTTGGGAGAACTCTCATCATAGCTATTTGGTCTTTTGA 524
 Qy
 729 AGCCATGTCTCTATTGAAATGATGGATCCCAAGATGATGCTGGCATGATTGGAACCA 788
 Db
 525 AGCCATGTCTCTATTGAAATGATGGATCCCAAGATGATGCTGGCATGATTGGAACCA 584
 Qy
 789 AGTTAATCGAAAGTTCTCAATTTTGAACAAGCTATCAAGATGCGCATTTAAATTA 848
 Db
 585 AGTTAATCGAAAGTTCTCAATTTTGAACAAGCTATCAAGATGCGCATTTAAATTA 644
 Qy
 849 AGATCTCACCTTGTGCTGAACGTAGGATATGATACATGTTTTCGCTGTTTCATAC 908
 Db
 645 AGATCTCACCTTGTGCTGAACGTAGGATATGATACATGTTTTCGCTGTTTCATAC 704
 Qy
 909 GTGTTTGAAGGCCATTCACTGGCACACAGATATTTACGTGCTTTTACATTCATATCC 968
 Db
 705 GTGTTTGAAGGCCATTCACTGGCACACAGATATTTACGTGCTTTTACATTCATATCC 764
 Qy
 969 AGACTTTAGAGATCTGCTATGAAGCTTTTGTCTGGGATCTTGAATCTGTGA 1028
 Db
 765 AGACTTTAGAGATCTGCTATGAAGCTTTTGTCTGGGATCTTGAATCTGTGA 824
 Qy
 1029 CATTGCAAGGAAAAAGTAAATGAAGCTGCTGTTTTCGAAGAGGAAGATTTTCAGTCAAT 1088
 Db
 825 CATTGCAAGGAAAAAGTAAATGAAGCTGCTGTTTTCGAAGAGGAAGATTTTCAGTCAAT 884
 Qy
 1089 GACTTATGGATTTAAATGGCTAAACAGTGTGACAGATCTTCAGTACAGGCATGCTAAA 1148

Db
 885 GACTTATGGATTTAAATGGCTAAACAGTGTGACAGATCTTCGAGTTACAGGCATGCTAAA 944
 Qy
 1149 AGATGTGGAGGATGACATGCAAGAAAGAGTAAGAGTACTCGAAGTCGACAAAGGAGAAGA 1208
 Db
 945 AGATGTGGAGGATGACATGCAAGAAAGAGTAAGAGTACTCGAAGTCGACAAAGGAGAAGA 1004
 Qy
 1209 AAGAGATCCAGAGCTTGAACCTAGAACACCAACCAATGTTTAGCAGTATTACAGCAGAGTCAA 1268
 Db
 1005 AAGAGATCCAGAGCTTGAACCTAGAACACCAACCAATGTTTAGCAGTATTACAGCAGAGTCAA 1064
 Qy
 1269 ATTACTCGTGTGTTACTGACAGTGCCTTATAGCCCTTTACTAAGAAAGACAGCAGTGTGT 1328
 Db
 1065 ATTACTCGTGTGTTACTGACAGTGCCTTATAGCCCTTTACTAAGAAAGACAGCAGTGTGT 1124
 Qy
 1329 TGCAGAGCTCAAAAATTTGATGTTCAAGCAGCAGATCTTCTTTCGTCATTCATAATTC 1388
 Db
 1125 TGCAGAGCTCAAAAATTTGATGTTCAAGCAGCAGATCTTCTTCTGTCATTCATAATTC 1184
 Qy
 1389 ATTGCATCATGGCATCCAGGCCAGCAATGATACATAAAAGAGATC 1435
 Db
 1185 ATTGCATCATGGCATCCAGGCCAGCAATGATACATAAAAGAGATC 1231
 RESULT 10
 LOCUS AX664135 719 bp DNA linear PAT 22-MAR-2003
 DEFINITION Sequence 7 from Patent WO02057304.
 ACCESSION AX664135
 VERSION AX664135.1 GI:29164145
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Panzer, S.R., Lincoln, S.E., Altus, C.M., Dufour, G.E., Jackson, J.L.,
 Jones, A.L., Dam, T.C., Liu, T.F., Harris, B., Flores, V., Daffo, A.,
 Marwaha, R., Chen, A.J., Chang, S.C., Gerstein, E.H., Peralta, C.H.,
 David, M.H. and Lewis, S.A.
 TITLE Secretary molecules
 JOURNAL Patent: WO 02057304-A 7 25-JUL-2002;
 Incyte Genomics, Inc. (US)
 FEATURES Location/Qualifiers
 source 1..719
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 /note="Incyte ID No: LI:462889.1:2001JAN12"
 ORIGIN
 Query Match 23.2%; Score 717.4; DB 6; Length 719;
 Best Local Similarity 99.9%; Pred. No. 8.4e-152;
 Matches 718; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 474 GGCGCGCGCGGAGCGCGCTGTTATTTTCCTGTGTCGACAGTGCCTGGCGCGCGCGGT 533
 Db
 1 GGCGCGCGCGGAGCGCGCTGTTATTTCCCTGTGTCGACAGTGCCTGGCGCGCGCGGT 60
 Qy
 534 GACCACGGGAGAGTAGGCATTAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGG 593
 Db
 61 GACCACGGGAGAGTAGGCATTAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGG 120
 Qy
 594 ATGGGAGCTCAGTATGCCAGAAAAATGGAGAAAAACATACAACTGGGTGGCAATTCAC 653
 Db
 121 ATGGGAGCTCAGTATGCCAGAAAAATGGAGAAAAACATACAACTGGGTGGCAATTCAC 180
 Qy
 654 CCAAGATTTTGAAGAAGCTTGTCCAGAAATTAAGTTGGGAAACTACTTTCATGATAAGCT 713
 Db
 181 CCAAGATTTTGAAGAAGCTTGTCCAGAAATTAAGTTGGGAAACTACTTTCATGATAAGCT 240
 Qy
 714 ATTTGGTCTTTTGAAGCCATGCTGCTATTGAAATGATGATCCCAAGATGGATGCTGG 773

```
Db 241 ATTGGTCCTTTTGAAGCCATGCTGCTATTGAAATGATGATCCCAAGATGATGCTGG 300
QY 774 CATGATGGAAACCAAGTTTAATCGAAAGTTCTCAATTTGACCAAGCTATCAAGATGG 833
Db 301 CATGATGGAAACCAAGTTTAATCGAAAGTTCTCAATTTGACCAAGCTATCAAGATGG 360
QY 834 CACTATTAATAATTAAGATCTCACTTGCCTGCACTGATGATGAGGATATGATACATGTTT 893
Db 361 CACTATTAATAATTAAGATCTCACTTGCCTGCACTGATGATGAGGATATGATACATGTTT 420
QY 894 TTGCTGTTGATAACGTTGTTAGAGGCCATTCACCTGGCACACAGATATTACGTGCCT 953
Db 421 TTGCTGTTGATAACGTTGTTAGAGGCCATTCACCTGGCACACAGATATTACGTGCCT 480
QY 954 TTACATTCATTAATCCAGACITTTATAGAGATCTGCTATGAAAGCTTTTCTCTCGGAAT 1013
Db 481 TTACATTCATTAATCCAGACITTTATAGAGATCTGCTATGAAAGCTTTTCTCTCGGAAT 540
QY 1014 CTTGAAAATCTGTGACATTCGAAGGAAAAGTAATAAAGCTGCTGTTTTCGAAGAGGA 1073
Db 541 CTTGAAAATCTGTGACATTCGAAGGAAAAGTAATAAAGCTGCTGTTTTCGAAGAGGA 600
QY 1074 AGATTTTCAGTCAATGCTTATGATTTAAATGGCTTAAAGTGTGACAGATCTTCAGT 1133
Db 601 AGATTTTCAGTCAATGCTTATGATTTAAATGGCTTAAAGTGTGACAGATCTTCAGT 660
QY 1134 TACAGGCATCTAAAGATGTGAGAGTGCATGCAAGAGAGTAAAGAGTCTCGAA 1192
Db 661 TACAGGCATCTAAAGATGTGAGAGTGCATGCAAGAGAGTAAAGAGTCTCGAA 719

RESULT 11
AC102010
LOCUS Mus musculus clone RP24-531M24, DNA linear HTG 21-MAR-2003
DEFINITION Mus musculus clone RP24-531M24, WORKING DRAFT SEQUENCE, 7 unordered
pieces.
ACCESSION AC102010
VERSION GI:29135693
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 147876)
Birren, B., Nusbaum, C. and Lander, E.
Mus musculus, clone RP24-531M24
Unpublished
2 (bases 1 to 147876)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepey, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
Lamaze, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL
REFERENCE
AUTHORS
```

REFERENCE
AUTHORS

```
3 (bases 1 to 147876)
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepey, Y.,
Collymore, A., Cook, A., Cooke, P., Corum, B., DeArelano, K.,
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (21-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 21, 2003 this sequence version replaced gi:28412092.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L17873
Center clone name: 531.M.24
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 146099 bases at least Q40
Consensus quality: 146869 bases at least Q30
Consensus quality: 147004 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 147276; sum-of-ctnigs
Quality coverage: 9.9 in Q20 bases; agarose-fp
Quality coverage: 10.6 in Q20 bases; sum-of-ctnigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 ctnigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the ctnigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 84958: contig of 84958 bp in length
* 84959 85058: gap of 100 bp
* 85059 87411: contig of 2353 bp in length
* 87412 87511: gap of 100 bp
* 87512 89933: contig of 2422 bp in length
* 89934 90033: gap of 100 bp
* 90034 94076: contig of 4043 bp in length
* 94077 94176: gap of 100 bp
* 94177 102536: contig of 8360 bp in length
* 102537 102636: gap of 100 bp
* 102637 118954: contig of 16318 bp in length
* 118955 119054: gap of 100 bp
* 119055 147876: contig of 28822 bp in length.
* Location/Qualifiers
* 1. 147876
* /organism="Mus musculus"
* /mol_type="genomic DNA"
```

TITLE
JOURNAL
COMMENT

```
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L17873
Center clone name: 531.M.24
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 146099 bases at least Q40
Consensus quality: 146869 bases at least Q30
Consensus quality: 147004 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 147276; sum-of-ctnigs
Quality coverage: 9.9 in Q20 bases; agarose-fp
Quality coverage: 10.6 in Q20 bases; sum-of-ctnigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 ctnigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the ctnigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 84958: contig of 84958 bp in length
* 84959 85058: gap of 100 bp
* 85059 87411: contig of 2353 bp in length
* 87412 87511: gap of 100 bp
* 87512 89933: contig of 2422 bp in length
* 89934 90033: gap of 100 bp
* 90034 94076: contig of 4043 bp in length
* 94077 94176: gap of 100 bp
* 94177 102536: contig of 8360 bp in length
* 102537 102636: gap of 100 bp
* 102637 118954: contig of 16318 bp in length
* 118955 119054: gap of 100 bp
* 119055 147876: contig of 28822 bp in length.
* Location/Qualifiers
* 1. 147876
* /organism="Mus musculus"
* /mol_type="genomic DNA"
```

FEATURES
source

```
1. 147876
/organism="Mus musculus"
/mol_type="genomic DNA"
```

```
/db xref="taxon:10090"
/clone="RP24-531M24"
/clone_lib="RPC1-24 Male Mouse BAC"
1. .84958
/clone="assembly_fragment
clone_end:SP6
vector_side:left"
85059. .87411
/note="assembly_fragment"
87512. .89933
/note="assembly_fragment"
90034. .94076
/note="assembly_fragment"
94177. .102536
/note="assembly_fragment"
102637. .118954
/note="assembly_fragment"
119055. .147876
/note="assembly_fragment
clone_end:T7
vector_side:right"

ORIGIN
Query Match 21.5%; Score 664.2; DB 2; Length 147876;
Best Local Similarity 72.6%; Pred. No. 2.3e-139;
Matches 1105; Conservative 0; Mismatches 293; Indels 125; Gaps 14;

Qy 1647 TGTCTTTCAAGATCTCTGTTACAAACAC- TTTCTGGTGGATACAAAGAGTCTTTG 1705
Db 72297 TGTCTTTCAAGATCTCTGTTACAAACAC TTTCTGGTGGATACAAAGAGTCTTTG 72356

Qy 1706 G-AACTCATCTCAGCAGACATGCTGAAGATGCATCTCGTCTTTTGTGAGTCTCTCC 1764
Db 72357 GAAACTCTCTCATGAAGACATGCTGAAGATGCATCTCGTCTTTTGTGAGTCTCTCC 72413

Qy 1765 GAGTGTCTTTCCCAAGTGTCTATATATAATCACCAGCTAAGGACTGTATCGACT 1824
Db 72414 CAGTGTCTATCCCTAAGTGTGCTATATATATGATCACCAGCTAAGGACTGTATCGACT 72473

Qy 1825 CTTTGTGATCACTGTGTTGGGCAATCTGTAGTCTTTATTCAGATCCTGACATACAA 1884
Db 72474 ACTTTGTACTCACTGTGTTGGGCAATCTGTAGTCTTTATTCAGATCCTGACATACAA 72533

Qy 1885 GGGCTCGACAGAGATGAAGCTTGTGATATCTTTGAGGATTTGCCACCTTGAGGATG 1944
Db 72534 GGGCTCGACAGAGATGAAGCTTGTGATATCTTTGAGGATTTGCCACCTTGAGGATG 72593

Qy 1945 AGTTTATGAC-----ATTTTATTTTAAATAGGCAGAGA 1976
Db 72594 AGTTTATGAC-----ATTTTATTTTAAATAGGCAGAGA 1976

Qy 1977 AGTTTATGACGCTTACACCATGCTGTTGAAACAGGAAACCCCAAGGCAACATTTGG 2036
Db 72654 ACTTTGATGAGTGTCTTACACTCGGTTGTTGAAACAGGAAACCCCAAGGCAACATTTGG 72713

Qy 2037 CTTGTTAGGTAACCTGCTGTTTACCATACCTTTCGATTTATGATACAGTACCTTCTAA 2096
Db 72714 CTTGTTAGGTAACCTGCTGTTTACCATACCTTTCGATTTATGATACAGTACCTTCTAA 72773

Qy 2097 GTGCTTTGAAATGGAACTCTACAGTATGACAGTACTATTAATATATTTGATCTCT 2156
Db 72774 TTGCTTTGAGTGGAGCTGACAGCAGCATGATTTCTACTCATGTTAGTGGTACCTCT 72833

Qy 2157 CTGAATTCCTTTAGCATGTTGATGTCACATTCAGTCTGCGGATGCTCTCAATGG 2216
Db 72834 CTGAATTCCTTTAGCATGTTGATGTCACATTCAGTCTGCGGATGCTCTCAATGG 72883

Qy 2217 CAGAGGAAAGGATAATGAAGAGAGCAGAGAAAGGCGGTAGTAG----- 2259
Db 72884 CAGAGGTAAGATAATGAGGAGCAGAGAAAGGAGCAGAGAAAGGCGGTAGTAG----- 72943

Qy 2260 -----TAAAAAACAAAGAAAAAAGAGTTCCGCCATTTGAGCCGAGAGATC 2308
Db 72943 -----TAAAAAACAAAGAAAAAAGAGTTCCGCCATTTGAGCCGAGAGATC 2308
```

```
72944 ACAAAAAAACAACAAAAAAGAAAGAAAGAGTTCTGTTGTTGAGCCAGAGATC 73003
Qy 2309 ACAATGAGCAAGCATATCAGAACATGTTGCTGCTGAATGTTTAAACCATGTTAGCATTT 2368
Db 73004 ATAATCAGCCAGGCTATCAGAA-ATGTTGGCTGGAATGTTTCAAGATGATGTTGATGTT 73062
Qy 2369 GACATGACCGCAAGTACGTTAAACCGAAGTTTGAAGTCTGATAGTGAACAGTTCGGTAT 2428
Db 73063 GACATGATGGCAAGTGTCTCAAGCCCAAGTTTGAAGTCTGATAGTGAACAGTTCGGTAT 73122
Qy 2429 GAACACAGGTTTGTCTCAATTCACACAGTGTGATG-----ACCCGCGCCAGTGCAC 2480
Db 73123 GAGCACAGGTTCTGTTTGTTCACACAGTGTGATGTTTACACTGCCACAGTGAAT 73182
Qy 2481 ACTTACAGTTCAAGGAAATGTTGACCTCAATAAATATAGCCCTCTCTCTCAGTCTCTG 2540
Db 73183 ACCTGCCGTTTCAAGGAAATGTTGACCTCAATAAATATAGCCCTCTCTCTCAGGCTCCAG 73242
Qy 2541 AACTGTATGTGGCAGCTAGTAAGCAGCTTTCAACAGGCAAAAATGATATTGAAAAATATTC 2600
Db 73243 AGCTTTATGTGGCAGCTAATAAGCAGCTTTCAACAGGCAAAAATGATCTCTGAAAAATGTC 73302
Qy 2601 CTAACCCGACCATGAGTTAATAAGATTTTAAAGTTTGCACCAACCACTTTGTGGTTA 2660
Db 73303 CCAATCCAGCAGTGAAGTCAAGTGAATCTTTAAGTTTGTAACTTCAACTTTCTGCTTA 73362
Qy 2661 TGAAGTTATTTGGCAGGAGGACACAAAAGGAATCTAAAGTTCTCTCTCTGAAATTTGATTTCT 2720
Db 73363 TGAAGCTCTTTGGCAGGAGGACACAAAAGAGAGTCAAAAGTTCTCTCTCCGAAATTTGACTTCT 73422
Qy 2721 CTGCTCATATAATATTTTCTGTTGTTGAAACTTTTTCAGAGAGACTGGGGAGTGGCCAT 2780
Db 73423 CTGTTCAATAATCTTTTCAATATTTGTAAGTGAATTTGAAGAGACTGCAAGAGCAGTGT 73482
Qy 2781 AAAGGGCAGAGTCTCTTTTTCAGACCCCACTTTTAGAGGGCAGATCACCAGGCTCCACAT 2840
Db 73483 GT-----CCACTCTAGGGGATACATCAGTCTGCCACAT 73515
Qy 2841 CACGGGAGTGAAGTGAATTTCTTGGGT-----ACCACTCATTTATAGGATACTTTT-A 2895
Db 73516 ATGTTGAGGTTAAAGCGGATTTCTTGAATGACAGACAGCTCTCTTAAAGAAATCTTTTCA 73575
Qy 2896 GTTTGACAGCTTATATGACATGAATGAAACTCTGTTTAAAGTGGTTTATTATGTTTC 2955
Db 73576 GTTTAAACCTTTACCTGGCATGTAAGTGAATGTAAGTGGTTTATTATG--- 73632
Qy 2956 CATGGAAGAACTGGTCTTATTTGAATGCAATGTAAGTGAATGTTTATTATGATGAT 3015
Db 73633 -----GGTGTAACTGGGATAGAAAGTATTTGGTGAATGTTTACATGATTTTATTATATAC 73685
Qy 3016 TTAAT--CACAAATCATTTTATGAATGATTTGAGTGAATAATAGTGTTTATAAGGTTAA 3073
Db 73686 TTAATGTTTAAACATTTTAAAGAAATGATTTGATG- AAATGGTGTGTTGTAAGAGTTAA 73744
Qy 3074 TAAATTTCTTGACAAAAAATAA 3096
Db 73745 TAAATTTCTTGACAAAAAATAA 73767

RESULT 12
AL353743
LOCUS Human DNA sequence from clone Rpl1-213G2 on chromosome 9, complete
DEFINITION AL353743
ACCESSION AL353743
VERSION AL353743.22 GI:15552931
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 181864)
AUTHORS Corby,N.
```

TITLE
JOURNAL

COMMENT

Direct Submission
Submitted (08-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Sep 11, 2001 this sequence version replaced gi:14267784.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30);
an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em.; EMBL; Sw.;
SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr9>
RP11-213G2 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-213G2. It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true right end of clone RP11-213G2 is at 181864 in this
sequence. The true left end of clone RP11-65C15 is at 111320 in
this sequence. The true right end of clone RP11-202111 is at 100 in
this sequence.

FEATURES

source
1..181864
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-213G2"
/clone_lib="RPCI-11.1"
39727

misc_feature
115957..116075
/notes="Sequence from overlapping clone BA65C15 (AL161453).
Assembly confirmed by restriction digest."
117488..117564
/notes="Sequence from overlapping clone BA65C15 (AL161453).
Assembly confirmed by restriction digest."

ORIGIN

Query Match 17.8%; Score 551.4; DB 9; Length 181864;
Best Local Similarity 99.8%; Pred. No. 9.4e-114;
Matches 552; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 TTCTCTCAGCAAACTCCAGCGCTGTATAGAAACATAAATCCGTTGTTCAGGCAGCAGTA 60
174760 TTCTCTCAGCAAACTCCAGCGCTGTATAGAAACATAAATCCGTTGTTCAGGCAGCAGTA 174819

61 GCACGCTGTGCTCTCGGAGCTGGCTGCTCGTTCGTCGTCGCAACCACTAAGCTCTACG 120
174820 GCACGCTGTGCTCTCGGAGCTGGCTGCTCGTTCGTCGTCGCAACCACTAAGCTCTACG 174879

121 CAAACCTCCACGGTTTCCTTCGCGCTTCGCGTCAACCTTCCTAGAAATCCCGAGAGGCA 180
174880 CAAACCTCCACGGTTTCCTTCGCGCTTCGCGTCAACCTTCCTAGAAATCCCGAGAGGCA 174939

181 GCCCAGACGGGGGGCTCTGAGACTCCGGGCTCCGCTCTTTCGGGGAACCGCCCACTA 240

Db 174940 GCCCAGACGGGGGGCTCTGAGACTCCGGGCTCCGCTCTTCCCGGAAACCGCCCACTA 174999

Qy 241 CCACAGACTCCGACAGAGGGTGAATAAAGATAAATCTCCGCTCTCGCGATCGTCTCTAATC 300
|||||

Db 175000 CCACAGACTCCGACAGAGGGTGAATAAAGATAAATCTCCGCTCTCGCGATCGTCTCTAATC 175059
|||||

Qy 301 TCAGCAGAGAGAGAGCGCGCCGATCGCGCCGAAACCGAGCGCTGCGAGGAGGGGGTG 360
|||||

Db 175060 TCAGCAGAGAGAGAGCGCGCGCCGATCGCGCCGAAACCGAGCGCTGCGAGGAGGGGGTG 175119
|||||

Qy 361 TGCCCGGGAGCGCGAAGTCCCGGGAGTAAGGAGAGGGGGCGGGTCCGCGTCCCGG 420
|||||

Db 175120 TGCCCGGGAGCGCGAAGTCCCGGGAGTAAGGAGAGGGGGCGGGTCCGCGTCCCGG 175179
|||||

Qy 421 GCATACGCATCGCTGCACGCTCGCGCTCGGGCTGAGAGGGGAGGGGGCGGGCGG 480
|||||

Db 175180 GCATACGCATCGCTGCACGCTCGCGCTCGGGCTGAGAGGGGAGGGGGCGGGCGG 175239
|||||

Qy 481 GCCGAGCGCGCTGCTGTTATTTCCGTCGTCGCGACAGTCGTCGCGGGCGGGGTGACACG 540
|||||

Db 175240 GCCGAGCGCGCTGCTGTTATTTCCGTCGTCGCGACAGTCGTCGCGGGCGGGGTGACACG 175299
|||||

Qy 541 GGAGAAAGTAGGCA 553
|||||

Db 175300 GGAGAAAGTAGGTA 175312
|||||

RESULT 13

LOCUS AX778899 440 bp DNA linear PAT 14-JUL-2003
DEFINITION Sequence 1056 from Patent WO03039443.
ACCESSION AX778899
VERSION AX778899.1 GI:32695893
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Haferlach,T., Schoch,C., Kern,W., Kohlmann,A., Schnittger,S.,
Dugas,M., Bils,R., Brors,B. and Mergenthaler,S.
TITLE Novel genetic markers for leukemias
JOURNAL Patent: WO 03039443-A 1056 15-MAY-2003;
Deutsches Krebsforschungszentrum (DE);
Ludwig-Maximilian-Universitaet Muenchen (DE);
PD Dr. Dr. (DE); Schoch, Claudia (DE); Kern, Wolfgang (DE)

FEATURES

Location/Qualifiers
1..440
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 14.2%; Score 440; DB 6; Length 440;
Best Local Similarity 100.0%; Pred. No. 7e-89;
Matches 440; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2548 TTGGCAGCTAGTAGCACTTTCAACAGGCAGCAAAATGATTCGAAATATATTCCTAACCC 2607
1 TTGGCAGCTAGTAGCACTTTCAACAGGCAGCAAAATGATTCGAAATATATTCCTAACCC 60

Db 1 TTGGCAGCTAGTAGCACTTTCAACAGGCAGCAAAATGATTCGAAATATATTCCTAACCC 2607

Qy 2608 GGACCATAGGTTAATAGAAATTTAAAGTTGCCAAACCAACTTTGTGGTTATGAAGTT 2667
61 GGACCATAGGTTAATAGAAATTTAAAGTTGCCAAACCAACTTTGTGGTTATGAAGTT 120

Qy 2668 ATTGGCAGAGACACAAAAGAAATCTAAGTTCTCTCTGAATTTGATTTCTCTGCTCA 2727
121 ATTGGCAGAGACACAAAAGAAATCTAAGTTCTCTCTGAATTTGATTTCTCTGCTCA 180

Qy 2728 TAAATATTTCTCTGTGTGAAACTTTGTTGAGAGAGACTGGGGAGTGCCCATAAAGGGG 2787
181 TAAATATTTCTCTGTGTGAAACTTTGTTGAGAGAGACTGGGGAGTGCCCATAAAGGGG 240

QY 2788 CAGAGTCCTCTTTTCAGACCCCAACTCTAGAGGGCACATCCAGGCTCCACATCACGGGA 2847
 Db 241 CAGAGTCCTCTTTTCAGACCCCAACTCTTAGAGGGCACATCCAGGCTCCACATCACGGGA 300
 QY 2848 AGTCAGATGGATTTCTTGGGTAAACAACCTATTATAAGGAATACATTTTAGTTTGACAGCCT 2907
 Db 301 AGTCAGATGGATTTCTTGGGTAAACAACCTATTATAAGGAATACATTTTAGTTTGACAGCCT 360
 QY 2908 TATATGACATGAATGAAGAACTGCTGTTTTAAAGTGGTTTATATATGTTCCATGGAAGAAAC 2967
 Db 361 TATATGACATGAATGAAGAACTGCTGTTTTAAAGTGGTTTATATATGTTCCATGGAAGAAAC 420
 QY 2968 TGGTCTTATTGAATGCATTG 2987
 Db 421 TGGTCTTATTGAATGCATTG 440

RESULT 14
 AL161447 160115 bp DNA linear PRI 16-MAR-2002
 LOCUS Human DNA sequence from clone RP11-379P1 on chromosome 9, complete
 DEFINITION sequence.
 ACCESSION AL161447
 VERSION AL161447.20 GI:19572746
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Johnson.C.

Direct Submission
 Submitted (16-MAR-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk
 On Mar 21, 2002 this sequence version replaced gi:15787718.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPep; Information on the WORMPEP
 database can be found at
<http://www.sanger.ac.uk/Projects/c-elegans/wormpep> This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
 Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr9>

RP11-379P1 is from the library RPECI-11.2 constructed by the group
 of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6
 This sequence is the entire insert of clone RP11-379P1. The true
 left end of clone RP11-280P22 is at 137145 in this sequence. The
 true right end of clone RP11-65C15 is at 38497 in this sequence.

FEATURES

Location/Qualifiers
 1..160115
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="9"
 /clone="RP11-379P1"
 /clone_lib="RPECI-11.2"

misc_feature complement(32505..32575)
 /note="Sequence from uni-directional dGTP big dye
 terminator reads only."
 74372..74537
 misc_feature /note="Single clone region. Assembly confirmed by
 restriction digest data."
 74706..74920
 misc_feature /note="Single clone region. Assembly confirmed by
 restriction digest data."
 complement(74706..74920)
 misc_feature /note="Sequence from uni-directional dGTP big dye
 terminator reads only."
 74921
 misc_feature /note="Tandem repeat. Forced join. Gap size estimated to
 be approximately 200bp by restriction digest data."
 129108
 misc_feature /note="Tandem repeat. Forced join. Gap size estimated to
 be approximately 300bp by restriction digest data."

ORIGIN

Query Match 12.9%; Score 398.4; DB 9; Length 160115;
 Best Local Similarity 99.8%; Pred. No. 4.9e-79;
 Matches 399; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2697 AAGTTCTCTCTGAATTTGATTTCTGCTCATATAATATTTTCCTGTTGTGAACTGTTT 2756
 Db 43136 AGTTCTCTCTGAATTTGATTTCTGCTCATATAATATTTTCCTGTTGTGAACTGTTT 43195
 QY 2757 GAGAGACATGGGAGGTGGCCATPAAGGGGCGAGAGTCTTCTTTTCAGACCAACTCTTAG 2816
 Db 43196 GAGAGACATGGGAGGTGGCCATPAAGGGGCGAGAGTCTTCTTTTCAGACCAACTCTTAG 43255
 QY 2817 AGGCGACATCACAGGCTCCACATCAGGGAGTGAGATGATTTCTTGGGTAACTC 2876
 Db 43256 AGGCGACATCACAGGCTCCACATCAGGGAGTGAGATGATTTCTTGGGTAACTC 43315
 QY 2877 ATTATAAGGAATATCTTTTAGTTTCAGAGCCTTATATGACATGAATGAATCTGTTT 2936
 Db 43316 ATTATAAGGAATATCTTTTAGTTTCAGAGCCTTATATGACATGAATGAATCTGTTT 43375
 QY 2937 AAAGTGTATTATATGTTTCCATGGAAGAACTCGTCTTATTTGAATGATGATGAACGTT 2996
 Db 43376 AAAGTGTATTATATGTTTCCATGGAAGAACTCGTCTTATTTGAATGATGATGAACGTT 43435
 QY 2997 ATATGTTTATTATACAGATTTAATCACAATCATTTTATGATGATGATGAATGAAT 3056
 Db 43436 ATATGTTTATTATACAGATTTAATCACAATCATTTTATGATGATGATGAATGAAT 43495
 QY 3057 GTGTTTATAAGGTTAATAAATTTCTTGACAAAAA 3096
 Db 43496 GTGTTTATAAGGTTAATAAATTTCTTGACAAAAA 43535

RESULT 15
 HSA334821/c 446 bp DNA linear PRI 18-JUL-2002
 LOCUS Homo sapiens genomic sequence surrounding NotI site, clone
 DEFINITION NBE-015R.
 ACCESSION AJ334821
 VERSION AJ334821.1 GI:15879239
 KEYWORDS
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 446)
 Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F., Kvasha,S.M.,
 Podewski,R.M., Matushkin,Y.G., Gyanchandani,A., Muravenko,O.V.,
 Levitsky,V.G., Kolchanov,N.A., Protopopov,A.I., Kashuba,V.I.,
 Kisselev,I.L., Wasserman,W., Wahlestedt,C. and Zabarovskiy,E.R.
 NotI flanking sequences: a tool for gene discovery and verification
 of the human genome
 Nucleic Acids Res. 30 (14), 3163-3170 (2002)

JOURNAL

MEDLINE 22131767
PUBMED 12136098
REFERENCE 2 (bases 1 to 446)
AUTHORS Zabarovsky, E.R.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre, Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77, Sweden

FEATURES
source Location/Qualifiers
1..446
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="NB6-015R"

ORIGIN

Query Match 10.4%; Score 321.4; DB 9; Length 446;
Best Local Similarity 99.7%; Pred. No. 5.8e-62;
Matches 322; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTCTCAGGAACTCCAGCGCTGTATAGGAAACATAAATCCGTTGTCTAGGCGAGCAGTA 60
Db |||||||
323 TTCTCAGGAACTCCAGCGCTGTATAGGAAACATAAATCCGTTGTCTAGGCGAGCAGTA 264
QY 61 GCACGCTGTTGCTCTCGAGCTTGGCTGCTCGTTTCGTTGCTCGCAACCACTAAGTCTTACG 120
Db |||||||
263 GCACGCTGTTGCTCTCGAGCTTGGCTGCTCGTTTCGTTGCTCGCAACCACTAAGTCTTACG 204
QY 121 CAAACCTCCACGGTTTCCTTCGCGCTTCGCGTCACTTTCTAAGAAATCCCGAGAGGCA 180
Db |||||||
203 CAAACCTCCACGGTTTCCTTCGCGCTTCGCGTCACTTTCTAAGAAATCCCGAGAGGCA 144
QY 181 GCGCAGACGGGGGGGTCTGAGACTCCGGGCTCCGGCTCTTTCCGGGAAACCGCCCACTA 240
Db |||||||
143 GCGCAGACGGGGGGGTCTGAGACTCCGGGCTCCGGCTCTTTCCGGGAAACCGCCCACTA 84
QY 241 CCCAGGACTCCGACAGAGGGTGAAGAAAGATAACTTCGGGTCTCGCGATCGTCTTAATC 300
Db |||||||
83 CCCAGGACTCCGACAGAGGGTGAAGAAAGATAACTTCGGGTCTCGCGATCGTCTTAATC 24
QY 301 TCGGAGAGAGAGAGCGCGCGC 323
Db |||||||
23 TCGGAGAGAGAGAGCGCGCGC 1

Search completed: August 11, 2004, 01:34:09
Job time : 12343.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 10, 2004, 17:10:23 ; Search time 8425.09 Seconds
(without alignments)
10973.574 Million cell updates/sec

Title: US-10-001-857-42

Perfect score: 3096

Sequence: 1 ttcttcacgaactccagg.....atttttgacaaaaaaa 3096

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_estba:*

3: em_estin:*

4: em_estin:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_esti:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_pig:*

27: em_gss_vil:*

28: gb_gssi:*

29: gb_gss2:*

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	795	25.7	1033	11	BC027201
2	777.8	25.1	973	11	AK021042
3	705	22.8	722	12	BM979022
4	696.8	22.5	854	10	BF969365

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

C	5	658.6	21.3	814	14	CB169246
	6	653.8	21.1	3071	11	AK050805
	7	653.6	21.1	768	12	BG547889
	8	648	20.9	957	13	BQ965085
	9	641.4	20.7	828	9	AI790514
	10	636.8	20.6	786	14	CD808079
	11	610	19.7	616	12	BM979511
	12	602.6	19.5	721	14	CF748928
	13	599.2	19.4	900	14	CA977646
	14	598.8	19.3	697	12	BI602827
	15	597	19.3	598	9	AW007583
	16	593.8	19.2	718	13	BU058931
	17	586	18.9	972	12	BG112812
	18	584.8	18.9	2958	11	BC043701
	19	575.2	18.6	763	13	BU613022
	20	572.6	18.5	723	14	CF725470
	21	572.4	18.5	952	13	BQ925670
	22	569.4	18.4	716	14	CF725998
	23	566	18.3	690	14	CF172281
	24	564.8	18.2	692	14	CF172718
	25	560.6	18.1	683	14	CF174274
	26	560	18.1	560	9	AA888870
	27	555.6	17.9	927	11	BC010475
	28	552.4	17.8	678	14	CF172404
	29	551.2	17.8	1117	13	BU525571
	30	550.2	17.8	671	14	CF902089
	31	549.6	17.8	672	14	CF172581
	32	548.4	17.7	873	9	AU081000
	33	542.2	17.5	663	14	CF916685
	34	542	17.5	735	13	BU057875
	35	540.6	17.5	663	14	CF916511
	36	540.6	17.5	684	14	CF724711
	37	540.2	17.4	661	14	CF904004
	38	539	17.4	663	14	CF913655
	39	538.6	17.4	654	10	BP468052
	40	538.2	17.4	660	14	CF173540
	41	538.2	17.4	660	14	CF914442
	42	537	17.3	537	12	BM699453
	43	526.8	17.0	666	13	BQ443173
	44	526	17.0	642	14	CF906238
	45	524.4	16.9	642	14	CF916163

ALIGNMENTS

RESULT 1	BC027201	1033 bp	mRNA	linear	HTC 19-NOV-2003
LOCUS	BC027201	Mus musculus	RIKEN cDNA C030004C14	gene, mRNA	(cdna clone)
DEFINITION	IMAGE:4483189)				containing frame-shift errors.
ACCESSION	BC027201				
VERSION	BC027201.1				GI:20070883
KEYWORDS	HTC.				
SOURCE	Mus musculus				(house mouse)
ORGANISM	Mus musculus				
REFERENCE	1				(bases 1 to 1033)
AUTHORS	Strausberg,R.L., Peingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Frange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,				

Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smallos, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22388257
 12477932
 2 (bases 1 to 1033)
 Direct Submission
 Strausberg, R.
 Submitted (04-APR-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbe-r@mail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amgobcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
 Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
 A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 31 Row: 1 Column: 19
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 21644570
 This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers
 1..1033
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4483183"
 /tissue_type="Mammary tumor. Metallothionien-TGF alpha
 model. 10 month old virgin mouse. Taken by biopsy."
 /clone_lib="NCI CGAP Mam1"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

ORIGIN

Query Match 25.7%; Score 795; DB 11; Length 1033;
 Best Local Similarity 88.3%; Pred. No. 3.3e-138;
 Matches 864; Conservative 0; Mismatches 115; Indels 0; Gaps 0;
 Qy 541 GGAGAGTAGGCATAATGGTTATCGAAGCTTCTGTAGATGATGACGATTACAGGTGGGAG 600
 Db |||||
 Qy 54 GCGCGGACGCATCATGGTTATGAAGCTTCTGTAGACGAGATGCTTCGGATGGGAG 113
 Db |||||
 Qy 601 CTCAGTATGCCAGAAAAAATGGAGAAAAACCAATACAACTGGGTGGACATTACCCAGAT 660
 Db |||||
 Qy 114 CTCGGGTCCCGAAAAAATGGAAAAAGTAGCACAACTGGGTGGACATAACCCAGAT 173
 Db |||||
 Qy 661 TTGAGAGAGCTGTGCGAAGTAATAAGTTGGAGAACTACTTTCATGATAGCTATTGGT 720
 Db |||||
 Qy 174 TTTGAAGATGCTGTGCGAGAGCTGAAGTTGGAGAACTGCTTCATGATAGCTGTTGGT 233
 Db |||||
 Qy 721 CTTTTTGAAGCCTGTCTGCTATTCAAAATGATGGAATCCCAAGATGGAGCTGGCATGATT 780
 Db |||||
 Qy 234 CTTTTTGAAGCCTGTCTGCTATTGAATGATGATGATGATGATGATGATGATGATGATG 293
 Db |||||
 Qy 781 GGAACCAAGTAAATCGAAAAATCTCTAATTTGAACCAAGCTATCAAGATGGCACTATT 840
 Db |||||

Db 294 GGGAAACCAAGTGAATAGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTT 353
 Qy 841 AAAATTAAAGATCTCACCCTTGCCTGAACTCATAGGAGTATAGGATACATGTTTTCGTGT 900
 Db 354 AAAATTAAAGACCTCAGCCTGCTGAACTCATAGGAGTATAGGATACATGTTTTCGTGT 413
 Qy 901 TTGATACAGTGGTGTAGAAAGCCCAATTCACCTGGCACAGACAGTATTTTACGTCCTTACATT 960
 Db 414 TTGATACAGTGGTGTAGAAAGCCCAATTCCTTGGCACAGACAGTATTTTACGTCCTTACATT 473
 Qy 961 CATAATCCAGACTTTTATAGAAAGTCTCTGATAGAGGCTTTTCTCTGGGAATCTTGAAT 1020
 Db 474 CATAATCCGACTTTCATAGAAAGTCTCTGATAGAGGCTTTTCTCTGGGAATCTTGAAT 533
 Qy 1021 ATCTGTACATTTGCAAGGGAAGTAATTAAGTCTCTGTTTGAAGAGGAGATTTT 1080
 Db 534 ATCTGTACATTTGCAAGGGAAGTAATTAAGTCTCTGTTTGAAGAGGAGATTTT 593
 Qy 1081 CAGTCAATGACTTTATGGATTTTAAATGGCTTAAACAGTGTGACAGATCTTTCAGTTTACAGGC 1140
 Db 594 CAGTCAATGACATACGATTTTAAATGGCTTAAACAGTGTGACAGATCTTTCAGTTTACAGGC 653
 Qy 1141 ATGCTAAAGAGTGTGGAGGATGACATGCAAGAGAGTAAGAGTACTCGAAGTCGACAA 1200
 Db 654 ATGCTAAAGAGTGTGGAGGATGATCTGCAAGAGGAGTAAAGAGTACTCGAAGTCGACAA 713
 Qy 1201 GGAGAAAGAGAGTCCAGAGTTGAACTAGACACCAACCAATCTTTAGCAGTATTCAGC 1260
 Db 714 GGAGAAAGAGAGATCCAGAGGTCGAATCAACACAGCAGTCTTGGCAGCATTCAGC 773
 Qy 1261 AGAGTGAATTTTACTCTGTGTACTGACAGTGTCTTATAGCCTTTTACTAAGAAAGAGAC 1320
 Db 774 AGAGTGAATTTTACTCTGTGTACTGACAGTGTCTTATAGCCTTTTACTAAGAAAGAGAC 833
 Qy 1321 AGTCTGTTGCAAGAGTCTTAAATTTGATGTTTCAAGCAGCAGATCTTCTTCTGCCATT 1380
 Db 834 AGTCTGTTGCAAGAGTCTTAAATTTGATGTTTCAAGCAGCAGATCTTCTTCTGCCATT 893
 Qy 1381 CATTAATTCATTGTCATATGTCATCCAGGCCAGATGATATCTACAAAGAGAGATCATCCA 1440
 Db 894 CACACCTTATGTCACACCGCATCCAGGCTCAGATGTCATGATGTCATCAAGAGAGATCATCCA 953
 Qy 1441 ATTATGATGGGTTTGAACCCCTTGTGAACCCAGAGGCTTCTTCCACCTTACCTTCCCTCGA 1500
 Db 954 ATTATGATGGGTTTGAACCCCTTGTGAACCCAGAGATTACTTCCACCCACCTTCCCTCGC 1013
 Qy 1501 TATGCAAAAATAAATAAAA 1519
 Db 1014 TATGCAAAAATAAATAAAA 1032

RESULT 2

AK021042

LOCUS

DEFINITION

AK021042

Mus musculus adult male corpus striatum cDNA, RIKEN full-length

enriched library, clone: C03004C14 product: CORNEAL WOUND HEALING

RELATED PROTEIN homolog [Rattus norvegicus], full insert sequence.

AK021042

Mus musculus

HTC; CAP trapper.

KEYWORDS

Mus musculus (house mouse)

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

99279253

10349636

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

QY 1381 CATAAATTCATGATCATGTCATCCAGCCAGAGTATGATACAAAGAGGATCATCCA 1440
 Db 849 CACACCTTCATTGACACCGCATCCAGGCTCAGAACTGGCACTACCAAGGAGACCATCCA 908
 QY 1441 ATTATGATGGGTTTGAACCCCTTGTGAACACGAGGCTACTTCCACCTACCTTCCCTCGA 1500
 Db 909 ATTATGATGGGTTTGAACCCCTTGTGAACACGAGATTAATCCACCCACCTTCCCTCGC 968
 QY 1501 TATGC 1505
 Db 969 TATGC 973

RESULT 3
 BM979022/c
 LOCUS
 DEFINITION 722 bp mRNA linear EST 21-FEB-2003
 UI-CF-DUI-adi-c-13-0-UI.s1 UI-CF-DUI Homo sapiens cDNA clone
 BM979022
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 8889548
 COMMENT
 Contact: McCray, PB
 McCray Lab
 University of Iowa
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
 Tel: 319 356 4866
 Fax: 319 356 7171
 Email: paul-mccray@uiowa.edu
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com) or from Open Biosystems
 (www.openbiosystems.com).
 Seq primer: M13 FORWARD
 POLYA=Yes.

FEATURES
 source
 1..722
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-CF-DUI-adi-c-13-0-UI"
 /tissue_type="Primary Lung Epithelial Cells"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-CF-DUI"
 /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 UI-CF-DUI is a normalized cDNA library containing the
 following tissue(s): Primary Lung Epithelial Cells. The
 library was constructed according to Bonaldo, Lennon and
 Soares, Genome Research, 6:791-806, 1996. First strand
 cDNA synthesis was primed with an oligo-dT primer
 containing a Not I site. Double stranded cDNA was ligated
 to an EcoR I adaptor, digested with Not I, and cloned
 directionally into pT7T3-Pac vector. The oligonucleotide
 used to prime the synthesis of first-strand cDNA contains
 a library tag sequence that is located between the Not I
 site and the (dT)18 tail. The sequence tag for this
 library is GGCTGTAGGC.

TAG TISSUE=Lung Epithelial Cells Tissue nos 359-368
 TAG_LIB=UI-CF-DUI
 TAG_SEQ=GGCTGTAGGC"

ORIGIN

Query Match 22.8%; Score 705; DB 12; Length 722;
 Best Local Similarity 99.9%; Pred. No. 2.2e-121;
 Matches 716; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 2380 CAAAGTACGTAAACCGAAGTTTGAGCTTGATAGTGAACAAGTTCGGTATGAACACAGGTT 2439
 Db 722 CAAAGTACGTAAACCGAAGTTTGAGCTTGATAGTGAACAAGTTCGGTATGAACACAGG-T 664
 QY 2440 TGCTCCATTCAACAGCTGTGATGACCCCGCCAGTGCACCTACTTACAGTTCAAGGAAT 2499
 Db 663 TGCTCCATTCAACAGCTGTGATGACCCCGCCAGTGCACCTACTTACAGTTCAAGGAAT 604
 QY 2500 GTCTGACCTCAATAAATATAGCCCTCTCTCAGTCTCTGAACTGTATGTGGCAGCTAG 2559
 Db 603 GTCTGACCTCAATAAATATAGCCCTCTCTCAGTCTCTGAACTGTATGTGGCAGCTAG 544
 QY 2560 TAAGCAGCTTCAACAGCCAAATAATGATTTGGAAATATTTCTAACCCGACCATGAGGT 2619
 Db 543 TAAGCAGCTTCAACAGCCAAATAATGATTTGGAAATATTTCTAACCCGACCATGAGGT 484
 QY 2620 TAATAGAAATTTAAAGGTTGCCAAACCAACTTTGTGTTATGAAGTTATTGGCAGGAGG 2679
 Db 483 TAATAGAAATTTAAAGGTTGCCAAACCAACTTTGTGTTATGAAGTTATTGGCAGGAGG 424
 QY 2680 ACACAAAAGGAATCTAAAGTTCCTCTGAAATTTGATTTCTCTCTCATATAATATTTTCC 2739
 Db 423 ACACAAAAGGAATCTAAAGTTCCTCTGAAATTTGATTTCTCTCTCATATAATATTTTCC 364
 QY 2740 TGTGTGAAACTGTTTTCAGAGAGACTGGGGAGGTGGCCATAAAGGGGACAGCTTCTT 2799
 Db 363 TGTGTGAAACTGTTTTCAGAGAGACTGGGGAGGTGGCCATAAAGGGGACAGCTTCTT 304
 QY 2800 TCAGACCCAACTCTTAGAGGCGCATCACACAGGCTCCACATCAGGGGAAGTGAGATGGAT 2859
 Db 303 TCAGACCCAACTCTTAGAGGCGCATCACACAGGCTCCACATCAGGGGAAGTGAGATGGAT 244
 QY 2860 TTCTTGGGTAACTCACTATTATAAGGAATCTTTTAGTTTTCAGAGCTTATATGACATGA 2919
 Db 243 TTCTTGGGTAACTCACTATTATAAGGAATCTTTTAGTTTTCAGAGCTTATATGACATGA 184
 QY 2920 ATGAAACTGCTGTTTAAAGTGGTTTATATGTTCCATGGAAGAACTGCTTATATGA 2979
 Db 183 ATGAAACTGCTGTTTAAAGTGGTTTATATGTTCCATGGAAGAACTGCTTATATGA 124
 QY 2980 ATGCATTGATGAAGCTTATATGTTTATATACAGATTAAATCACAATAATCATTTTATGA 3039
 Db 123 ATGCATTGATGAAGCTTATATGTTTATATACAGATTAAATCACAATAATCATTTTATGA 64
 QY 3040 ATGATTGAGTGAAAATAGTGTGTATAAGGTTAATAAATTTCTTGACAAAAA 3096
 Db 63 ATGATTGAGTGAAAATAGTGTGTATAAGGTTAATAAATTTCTTGACAAAAA 7

RESULT 4
 BF969365
 LOCUS
 DEFINITION BF969365 854 bp mRNA linear EST 22-JAN-2001
 602271630F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4359702 5',
 mRNA sequence.
 ACCESSION BF969365
 VERSION BF969365.1 GI:12336580
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 854)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
COMMENT

Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LENL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LENL at:
<http://image.llnl.gov>
 Plate: LLAM9999 row: 1 column: 07
 High quality sequence stop: 785.
 Location/Qualifiers
 1..854
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4359702"
 /tissue_type="adrenal cortex carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_84"
 /note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.229 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 22.5%; Score 696.8; DB 10; Length 854;
 Best Local Similarity 94.5%; Pred. No. 7.5e-120;
 Matches 809; Conservative 0; Mismatches 37; Indels 10; Gaps 8;

QY 2245 GAAAGCCGTAGTAGTAAAGAAACAAAG-AAAAAAGAAAGTTGCGCCATTGAGCGGAG 2303
 Db 1 GAAAGCCGTAGTAGTAAAGAAACAAAG-AAAAAAGAAAGTTGCGCCATTGAGCGGAG 60

QY 2304 AGATCAATGAGCGCAAGCATATCAGAACATGTGCTGGAATGTTTAAACCATGTAG 2363
 Db 61 AGATCAATGAGCGCAAGCATATCAGAACATGTGCTGGAATGTTTAAACCATGTAG 120

QY 2364 CATTTGACATGAGCGCAAGTAGTAAACCGAAGTTTGAGCTTGATAGTGAACAAAGTTC 2423
 Db 121 CATTTGACATGAGCGCAAGTAGTAAACCGAAGTTTGAGCTTGATAGTGAACAAAGTTC 180

QY 2424 GGTATGAACACAGTTTGTCTCCATTCAACAGTGTATGACCCCGCGCCAGTGCACTACT 2483
 Db 181 GGTATGAACACAGTTTGTCTCCATTCAACAGTGTATGACCCCGCGCCAGTGCACTACT 240

QY 2484 TACAGTTCAAGGAAATGTCTGACCTCAATAAATATAGCCCTCTCTCAGTCTCCTGAAC 2543
 Db 241 TACAGTTCAAGGAAATGTCTGACCTCAATAAATATAGCCCTCTCTCAGTCTCCTGAAC 300

QY 2544 TGTATGCGAGCTAGTAAGCACTTTCAACAGGCAAAATATGATTTGGAATAATTCCTA 2603
 Db 301 TGTATGCGAGCTAGTAAGCACTTTCAACAGGCAAAATATGATTTGGAATAATTCCTA 360

QY 2604 ACCCGACCATAGGTTAATAGATTTTAAAGTTGCGCAACCAACTTTGCGTTATGA 2663
 Db 361 ACCCGACCATAGGTTAATAGATTTTAAAGTTGCGCAACCAACTTTGCGTTATGA 420

QY 2664 AGTTATTGCGAGGAGACACAAAGGAATCTAAAGTTCTCTGTAATTTGATTTCTCTG 2723
 Db 421 AGTTATTGCGAGGAGACACAAAGGAATCTAAAGTTCTCTGTAATTTGATTTCTCTG 480

QY 2724 CTCATAAATATTTCTGTTGTGAACCTTTTGAAGAGACTGGGAGGTGGCCATAA 2783
 Db 481 CTCATAAATATTTCTGTTGTGAACCTTTTGAAGAGACTGGGAGGTGGCCATAA 540

QY 2784 GGGCGAGAGTCTTTTCAGACCAACTCTTAGAGGCGCATCAGCAGGCTCCACATCAC 2843
 Db 541 GGGCGAGAGTCTTTTCAGACCAACTCTTAGAGGCGCATCAGCAGGCTCCACATCAC 600

QY 2844 GGGAAGTGAGATGGATTTCTTGGTAACTCAATATAGGAATACTTTTAG-TTTGAC 2902

Db 601 GGAAGTGAGATGGATTTCTTGGTAACAACACTCATTATAAGGAATACTTTAGTTTGGAC 660
 QY 2903 AGCCTTATATGACA-TGAATGAAAACTGCTGTTTTAA-AGTGGTTTATTATGTTCCAT-G 2959
 Db 661 AGCCTTATATGACATTTGAATGAAAACTGCTGTTTTAAACAGTGGTTCTTATGTTCCCTGG 720
 QY 2960 GAAGAACTGCTCTTATTGAATGCAATGATGAACGTTATGTTTATTACAGATTAA 3019
 Db 721 GACGAACTGCTCTTATTGAATGCAATGATGAACG-TATATGGGTTTATTACCGATTAA- 778
 QY 3020 TCACAATCATTTTATTGAATGATGAGTGAATAGTCTTTATATAAGCTTTAATAAATT 3079
 Db 779 --TCCCAATCATTTTATTGAATGATG-TGGCAGATCGGTTTAAACGTGATAAATCTTGTG 835
 QY 3080 TCTTGACAAAAA 3095
 Db 836 CGCAAAAAACAAAAA 851

RESULT 5
 CBI69246/c
 LOCUS
 DEFINITION
 CBI69246
 ACCESSION
 VERSION
 CBI69246.1 GI:28155372
 SOURCE
 EST.
 Bos taurus (cow)
 ORGANISM
 Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 814)
 Adelson,D.L. and Gill,C.A.
 Bovine ESTs (Adelson and Gill)
 Unpublished (2003)
 Contact: David L. Adelson
 Animal Breeding and Genetics
 Texas A&M University
 Animal Science Dept., TAMU-2471, College Station, TX 77843-2471,
 USA
 Tel: 9798452616
 Fax: 9798456970
 Email: david.adelson@tamu.edu.
 Location/Qualifiers
 1..814
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /tissue_type="testes"
 /clone_lib="CSEQFN41 testes"
 /note="Organ: testes; Vector: pBluescript SK+; Site_1:
 NotI; Site_2: EcoRI; sequence 5' of the insert
 (5'-NNN...NNNinsert)
 GGAATTGAGCTCCACCGGTCGCGCGCGGCTCGAG. Sequence 3' of
 the inserts (AAGATTCGATCAAGCTTATGATACCTGACCTCGAG.
 normalized Rd 1 library, sequenced 3' with M13R primer."

FEATURES
 source
 1..814

ORIGIN
 Query Match 21.3%; Score 658.6; DB 14; Length 814;
 Best Local Similarity 90.3%; Pred. No. 1e-112;
 Matches 758; Conservative 0; Mismatches 54; Indels 27; Gaps 4;

QY 1348 ATGTTTCAAGCAGCAGATCTTCTTCTGCCATTCAATTAATTCATGTCATGTCATCCAG 1407
 Db 814 ATGTTTCAAGCAGCAGATCTTCTTCTGCCATTCAATTAATTCATGTCATGTCATCCAG 755
 QY 1408 GCCAGATGATCT-ACAAAAGGAGATCATCAATATGATGGGTTTGAACCCCTTGT 1466
 Db 754 GCCAGATGATCACTAACAAAAGGAGCATCCCATATGATGGGTTTGAACCACTTGT 695
 QY 1467 GAACAGAGGCTACTTCCACCTACCTTCCCTCGATATGCAAAATTAATTAAGGGAAGA 1526

[illegible]

Best Local Similarity 88.1%; Pred. No. 7.2e-112; Matches 736; Conservative 0; Mismatches 92; Indels 7; Gaps 2;									
QY	1316	AGACCAAGTCTGTTGAGAGCTCAAAATATGATGGTTCAGGAGAGATCTCTCTTTCTG	1375						
Db	210	AGACCAAGTCTGTTGAGAGCTCAGAAATGATGGTCCAGGCGGAGACCTCTTTCTG	269						
QY	1376	CCATTCAATATTCATTCATCATGGATCCAGGCCAGAGATGATCTACAAAAGGAGATC	1435						
Db	270	CCATTCACTCATTCGACACCGCATCCAGGCTCAGAATGGCACTACCAAAAGGAGACC	329						
QY	1436	ATCCAAATATGATGGTGTGAAACCCCTTGTGAACACAGAGGCTACTTCCACCTACCTTCC	1495						
Db	330	ATCCAAATATGATGGTGTGAGCCCTTGTGAACACAGAGATTAATTCACCCACCTTCC	389						
QY	1496	CTCGATATGCAAAATTAATTAAGGGAAGAAATGGTGAACTATTTTGAAGATTAATAG	1555						
Db	390	CTCGATATGCAAAATTAATTAAGGGAAGAAATGGTGAACTATTTTGAAGATTAATAG	449						
QY	1556	ATAGAATAAAATCTGCTGAGGTGCTGAATTTAAATAATTTACATTTGATCTCGATT	1615						
Db	450	ACAGAATAAAATCTGCTGAGGTGCTGAATTTAAATAATTTACATTTGATCTCGATT	509						
QY	1616	TTTCTGTGAATTTAGTGAACAGCTACCATGTGTCTTTCAAGATCTCTGTTACAAACCA	1675						
Db	510	TTTCTGTGAATTTAGTGAACAGCTACCATGTGTCTTTCAAGATCTCTGTTACAAACCA	569						
QY	1676	CTTCTCTGGTGAATAACAAAAGGCTTTGGAACTCATCTATCTGCAAGACATGGTGAAG	1735						
Db	570	CTTCTCTGGTGAATAACAAAAGGCTTTGGCACTCATCTATCTGCAAGACATGGTGAAG	629						
QY	1736	ATGCACCTCGGTCTTTGTGATGATCTCCGAGTCTTCCCAAGTGTACCTATATA	1795						
Db	630	ATGCTCTGGTCTTCTGTCAG--TCTCTCGTGTCTCTCCCAAGTGTGCTATATA	687						
QY	1796	TATATCCAGGCTAAGAGCTGATGACATCTCTTTGTACTCACTGTTGGCGCAATCTG	1855						
Db	688	TATATCCAGGCTAAGAGCTGATGACATCTCTTTGTACTCACTGTTGGCGCGTTTG	747						
QY	1856	TAGTCTTATTCAGATCCATGACATACAGGCTTCAGAGAGATAGCTTGGTCAAT	1915						
Db	748	TAGTCTTGTTCAGATCCATGGCCATACAGGCTTCGAGAGACAAGCTTGGTCAAT	807						
QY	1916	TCTTGAGGAATTTGCCACCTTGACAGTGTATGACATTTTATTAATAGGAGAG	1975						
Db	808	TCTTGAGGATTTGCTACCTTGACAGATGAT-----ACAATGTTTGAACAGGAGAG	862						
QY	1976	AAGTTTGTACAGCGTTTCAACCATGCTGTGAAACAGGAACCCCAAGGCAACATTTG	2035						
Db	863	AAGTTTGTACAGCGTTTCACTATATGTTTGAACAGGAGCTTCAGAGACAACTG	922						
QY	2036	GCCTGTTTAGGTACCTGGTCTTTTACCAATACCTTGGATATGATGATGATCTCTA	2095						
Db	923	GCCTGTTTAGGAACCTGGTCTTTTACCAATACCTCCGATATGATGATGATCTCTC	982						
QY	2096	AGTGGCTTTGAATTTGAACTTACAGTATGACAGTACTATTACATATATGCT	2150						
Db	983	AGTGGCTTTGAGCTGAGCTGTACAGATGATGATGATGATGATGATGATGATGAT	1037						
RESULT 7									
BG547889									
LOCUS									
DEFINITION	602576187F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4704212 5',								
	mRNA sequence.								
ACCESSION	BG547889								
VERSION	BG547889.1 GI:113546554								
KEYWORDS	EST.								
SOURCE	Homo sapiens (human)								
ORGANISM	Homo sapiens								
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;								
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.								
REFERENCE	1 (bases 1 to 768)								

AUTHORS
TITLE
JOURNAL
COMMENT

NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLCM1543 row: n column: 21
High quality sequence stop: 738.

FEATURES
source

1..768
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4704212"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH MGC 77"
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggcgctggcc); Site 2: SfiI (ggcattatggc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGGCGCATATGCCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGCGCGACATG-dn(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."

ORIGIN

Query Match		21.1%; Score 653.6; DB 12; Length 768;	
Best Local Similarity		96.9%; Pred. No. 8.8e-112;	
Matches		698; Conservative 0; Mismatches 19; Indels 3; Gaps 3;	
QY	2149	GTATCTCTCTGAATTCCTTTAGCATGGTTGATGTCAAATGAGTCGTCGCGATGGCTC	2208
Db	1	GTATCTCTCTGAATTCCTTTAGCATGGTTGATGTCAAATGAGTCGTCGCGATGGCTC	60
QY	2209	TCAAATGGCAGAGGAAAGGATTAATGAGCAGCAGAGAGGCGCTAGTAGTAAAAAAC	2268
Db	61	TCAAATGGCAGAGGAAAGGATTAATGAGCAGCAGAGAGGCGCTAGTAGTAAAAAAC	120
QY	2269	AAAGAAAAAAGAAAGTTGCGCCATTGAGCCGAGAGATCAATGAGCAAGCATATCA	2328
Db	121	AAAGAAAAAAGAAAGTTGCGCCATTGAGCCGAGAGATCAATGAGCCTAGCATATCA	180
QY	2329	GAACTGTGTCTGGAATGTTTAAACCATGGTAGCATTTGACATGGACGCGCAAGTACG	2388
Db	181	GAACTGTGTCTGGAATGTTTAAACCATGGTAGCATTTGACATGGACGCGCAAGTACG	240
QY	2389	TAAACCGAAGTTGAGCTTGATGTAAGTTCGATGTAAGTTCGATGTAAGTTCGATGTA	2448
Db	241	TAAACCGAAGTTGAGCTTGATGTAAGTTCGATGTAAGTTCGATGTAAGTTCGATGTA	300
QY	2449	CAACAGTGTGATGACCCCGCCGAGTGCATCTACTTACAGTTTCAAGGAATGCTGACCT	2508
Db	301	CAACAGTGTGATGACCCCGCCGAGTGCATCTACTTACAGTTTCAAGGAATGCTGACCT	360
QY	2509	CAATTA-ATATAGCCTCTCTCTAGTCTCTGAACTGTATGTCAGCTAGTAGGACT	2567
Db	361	CAATTAACATATAGCCTCTCTCTAGTCTCTGAACTGTATGTCAGCTAGTAGGACT	420
QY	2568	TTCAACAGGCAAAATGATATTGGAATATTCCTAACCCGACCATGAGGTTATAGAA	2627
Db	421	TTCAACAGGCAAAATGATATTGGAATATTCCTAACCCGACCATGAGGTTATAGAA	480
QY	2628	TTTTAAAGGTTGCCAAACCCAACTTTTGTGTTATGAAGTTATTGGCAGAGGACACAAA	2687
Db	481	TTTTAAAGGTTGCCAAACCCAACTTTTGTGTTATGAAGTTATTGGCAGAGGACACAAA	540

MGI:994033
Possible reversed clone: similarity on wrong strand
Seq primer: custom primer used
High quality sequence stop: 509.
Location/Qualifiers
1. .828
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:2064845"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Sugano mouse kidney mkia"
/note="Organ: kidney; Vector: pME18S-FL3; Site 1: DralIII (CACTGTGTG); Site 2: DralII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DralIII adaptor [TGTGGCCTACTGG], digested and cloned into distinct dralIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTAAAGCTGGC and 3' end primer CGACCTGCAGCTCGACACA."

ORIGIN

Query Match 20.7%; Score 641.4; DB 9; Length 828;
Best Local Similarity 86.4%; Pred. No. 1.7e-109;
Matches 716; Conservative 0; Mismatches 112; Indels 1; Gaps 1;

QY 673 TGTGCGAATAAAGTTGGGAGAACTACTTCATGATAAGCTATTGTCTTTTGAAGCC 732
DB 828 TGTGCGAGAGCTGAAGTTGGAAGACTGCTTCATTGATAAGCTGTTTCTTTTGAAGCC 769

QY 733 ATGCTCTGCTATTCAATATGATGGATCCAGATGGATCTGGGATGATTTGGAACCAAGTT 792
DB 768 ATGCTCTGCTATTCAATATGATGGATTCNTAAGATGGATCGGGTATGATCNGAACCAAGTG 709

QY 793 AATCGAAAAGTTCTCAATTTTGAACAAGCTATCAAGATGGCACTATTAATAATTAAGAT 852
DB 708 AATAGNAAAGTTCTCAATTTTGANCAAGCTATCAGATGGCAC-ATTAATTTAAAGAC 650

QY 853 CTCACCTGCTGAACTGATAGGATTTATGGATACATGTTTTCGTGTTTGATAAGCTGG 912
DB 649 CTCAGGCTGCTGAACTGATAGGATTAATAGACACCTGTTTCTGCTGTTTGATCAGATGG 590

QY 913 TTAGAGGCGCATTCACCTGGCAGACAGATATTTACGTGCTTTACATTCATATCCAGAC 972
DB 589 CTCGAAGGCGCATTCCTTTGGCAGACAGATGTTTACGTGCTTTACATTCATATCCCGAC 530

QY 973 TTTATAGAAGATCCTGCTATGAAGGCTTTTGTCTCTGGATCTTGAATCTGTGACATT 1032
DB 529 TTTATAGAAGATCCTGCTATGAAGGCTTTTGTCTCTGGATCTTGAATCTGTGACATT 470

QY 1033 GCAAGGGAAGATTAATAAAGCTGCTGTTTGAAGAGAGGATTTTCAGTCAATGACT 1092
DB 469 GCACGGGAAAAAGTAAATAAAGCTGCTGTTTGAAGAGAGGATTTTCAGTCAATGACA 410

QY 1093 TATGGATTTAAATGCTTAACAGTGTGCAGATCTTCAGATTCACGGATGCTTAAAGAT 1152
DB 409 TACGGATTTAAATGCTTAACAGTGTGCAGATCTTCAGATTCACGGATGCTTAAAGAT 350

QY 1153 GTGGAGGATGACATGCAAGAGAGTAAAGAGTACTCGAAGTCGACAGGAGAGAGA 1212
DB 349 GTGGAGATGATCTGCAAGAGGAGTAAAGAGTACTCGAAGTCGACAGGAGAGAGA 290

QY 1213 GATCCAGAAGTTGAATAGAACACCAACAAATGTTTACAGATTTACAGCAGAGTGAATTT 1272
DB 289 GATCCAGAAGTTCGAATAGAACCCACAGCAGTGTCTTGGCAGCATTCAGCAGAGTGAATTC 230

QY 1273 ACTGCTGTGTACTGACAGTGTCTTATAGCCCTTTACTAAGAAAGACAGAGTGTGTGCA 1332
DB 229 ACGGAGTGTCTGCTCAGTGTCTCATAGCCCTTTTACTAAGAAAGACAGTGTGTGCA 170

QY 1333 GAAGCTCAAAAATGTATGTTTCAAGCAGAGATCTTCTTCTGCTCATTCATATTCATTG 1392
DB 169 GAGGCTCAGAAACTGATGTTCCAGGGCGGACAGCTTCTTCTGCTCATTCACACCTATTG 110

QY 1393 CATCATGGCATCCAGGCCAGAAATGATACATAAAAGAGAGATCATCCAAATTCATGCGGT 1452
DB 109 CACCACGGCATCAGGCTCCGATGGCATCCAAAGAGAGACATCCAAATTAATGGGT 50

QY 1453 TTTGAAACCCCTTGTGAACACAGAGCTACTTCCACCTACCTTCCCTCGAT 1501
DB 49 TTTGAGCCCTTGTGAACACAGAGATTTTCCCCCGCTTCCCTCGAT 1

RESULT 10
CD808079
LOCUS
DEFINITION
UI-M-GW0-ciq-n-20-0-UI.r1 NIH BMAP_GW0 Mus musculus CDNA clone
IMAGE:30540139 5', mRNA sequence.
CD808079
ACCESSION
CD808079.1 GI:32466905
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 786)
NIH-MGC http://imgc.ncbi.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mouseffi.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers
1. .786
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30540139"
/tissue_type="whole eye"
/dev_stage="embryo.15.5,16.5,17.5,18.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP_GW0"
/note="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is CCGGCTCTC. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."

FEATURES
source

ORIGIN


```

Db      376  A T A A T A T T T T C C T G T T G T G A A C T T G T T T G A G A G A C T G G G A G G T G G C C A T A A A G G G   317
QY      2787  G C A G A G T C T T C T T T C A G A C C C A A C T T T A G A G G C A C A T C A C C A G G T C C A C A T C A C G G G   2846
Db      316  G C A G A G T C T T C T T T C A G A C C C A A C T T T A G A G G C C A C A T C A C C A G G T C C A C A T C A C G G G   257
QY      2847  A A G T G A G A T G A T T C T T G G T A C A A C T C A T T A T A A G G A T A C T T T T A G T T T G A C A G C C   2906
Db      256  A A G T G A G A T G A T T C T T G G T A C A A C T C A T T A T A A G G A T A C T T T T A G T T T G A C A G C C   197
QY      2907  T T A T A T G A C A T G A T G A A A A C T G C T G T T T T A A A G T G C T T T A T T A T T G T T C C A T G A A G A A   2966
Db      196  T T A T A T G A C A T G A T G A A A A C T G C T G T T T A A A G T G C T T T A T T A T T G T T C C A T G A A G A A   137
QY      2967  C T G G T C T T A T T G A T G A C A T G A A C G T T A T A T A T G T T T T A T T A C A G A T T T A A T C A C A A   3026
Db      136  C T G G T C T T A T T G A T G A C A T G A A C G T T A T A T A T A T A T A T A T A T A T A T T T C T T G A C   77
QY      3027  T C A T T T T T A T G A T T A T G A T G A A A A T A G T G T T T A T A A A G G T T A T A A A T T C T T G A C   3086
Db      76  T C A T T T T T A T G A T T A T G A T G A A A A T A G T G T T T A T A A A G G T T A T A A A T T T C T T G A C   17
QY      3087  A A A A A A A A A A   3096
Db      16  A A A A A A A A A A   7

```

```

RESULT 12
CF748928
LOCUS
DEFINITION
  CF748928 721 bp mRNA linear EST 10-OCT-2003
  UI-M-HJ0-cmi-1-06-0-UI.r1 NIH_BMAP_HJ0 Mus musculus cDNA clone
  IMAGE:30626477 5', mRNA sequence.
ACCESSION
  CF748928
VERSION
  EST.
KEYWORDS
  Mus musculus (house mouse)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 721)
REFERENCE
  NIH-MGC http://mgi.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: c9abps@mail.nih.gov
  Tissue Procurement: Dr. James Lin University of Iowa
  cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
  cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: Distribution information can be found at
  http://genome.uiowa.edu/distribution/mouseefi.html
  This clone was contributed by the Brain Molecular Anatomy Project
  (BMAP)
Seq primer: pYX-5.

```

```

FEATURES
  source
    1..721
    /organism="Mus musculus"
    /mol_type="mRNA"
    /strain="C57BL/6"
    /db_xref="taxon:10090"
    /clone="IMAGE:30626477"
    /tissue_type="Upper Head"
    /dev_stage="9.5 and 10.5 dpc"
    /lab_host="PH10B (T1 phage resistant)"
    /clone_lib="NIH BMAP HJ0"
    /note="Organ: Head; Vector: pYX-Asc; Site 1: ECOR I;
    Site 2: Not 1; The library was constructed according
    Bonafide, Lennon and Soares, Genome Research, 6:791-806,
    1996. Denatured RNA was size fractionated on a 1% agarose
    gel. First strand cDNA synthesis was primed with oligo-dT
    primer containing a Not I site. Double strand cDNA was
    size selected according to mRNA size fraction, ligated

```

with EcoR I adaptor , digested with NotI and then cloned directionally into pYX-Asc vector . The library tag sequence located between the Not I site and the polyA tail is CGAATCAAT. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

ORIGIN

```

Query Match      19.5%; Score 602.6; DB 14; Length 721;
Best Local Similarity 89.7%; Pred. No. 3e-102;
Matches 647; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 691  G G A A C T A C T T C A T G A T A G C T A T T T G T C T T T T T G A A G C C A T G T C T G T A T T G A A T G   750
Db      1  G G A A A C T G C T T C A T G A T A G C T G T T T G T C T T T T T G A A G C C A T G T C T G T A T T G A A T G   60
QY 751  A T G A T C C C A A G A T G A T G C T G C A T T T G G A A C C A A G T T A A T C G A A A G T T C T C A A T   810
Db      61  A T G G A T C C T A A G A T G A T G C C G T A T G A T C G G A C C A A G T A A T A G A A A G T T C T C A A T   120
QY 811  T T T G A A C A A G C T A T C A A G G A T G C A C T A T T A A A A T T A A A G A T C T C A C C T T G C C T G A A C T G   870
Db      121  T T T G A C A A G C T A T C A A G G A T G C A C A T T A A A A T T A A A G A C C T C A G C C T G C C T G A A C T G   180
QY 871  A T A G G A T T A T G G A T A C A T G T T T T T G T C T G T T T G A T A A C G T G T T A G A A G C C A T T C A C T G   930
Db      181  A T A G G A A T A A T G A C A C C T G T T T C T G T T G T T G A T C A C A T G G C T C G A G C C A T T C C T T G   240
QY 931  G C A C A C A G A T A T T A G T G C C T T T A C A T T C A T A T T C C A G A C T T T A T A G A A G A C C T G C T   990
Db      241  G C A C A C A G A T G T T A G T G C C T T T A C A T T C A T A A T C C G A C T T C A T A G A A G A T C C T G C C   300
QY 991  A T G A A G C T T T T G C T C T G G G A T C T T G A A A T C T G T G A C A T T C G A G G G A A A A G T A A A T   1050
Db      301  A T G A A A G C T T T T G C T C T G G A T C T T G A A G A T C T G C A C A T T C G A C C A T T C A C G G G A A A A G T A A A T   360
QY 1051  A A A G C T G C T G T T T T T T G A A G A G G A A G A T T T C A G T C A A T C A C T T A T G G A T T T A A A A T G G C T   1110
Db      361  A A A G C T G C T G T T T T T G A A G A G G A A G A T T T C C A G T C A A T G A C A T A C G G A T T T A A A A T G G C C   420
QY 1111  A A C A G T G T C A C A G A T C T T C A G T T A C A G G A T G C T T A A A A G A T G T G A G A G A T G A C A T G C A A   1170
Db      421  A A C A G T G T C A C A G A T C T C G A G T T A C A G G A T G C T A A A G A T G T G A A G A T G A T G A T C T G C A A   480
QY 1171  A G A A G A G T A A A G A G T A C T C G A A G T C A C A A G A G A G A A A A G A G A T C C A G A A G T T G A A C T A   1230
Db      481  A G C G A G T A A A G A G T A C T C G A A G T C G A A G A G A G A G A G A G A G A T C C A G A G G T C G A A C T A   540
QY 1231  G A A C C A A C A A T G T T T A G C A G T A T T C A G C A G A T G A A A T T T A C T C G T G T G T T A C T G A C A   1290
Db      541  G A A C A C C A G A C T G C T T G C A G A C A T T C A G C A G A T G A A G T T C A C G C G A G T G C T G C T C A C A   600
QY 1291  G T G C T T A T A G C C T T T A C T A A G A A G A G A C C A C T G C T G T T G C A G A G A C T C A A A A A T T G A T G   1350
Db      601  G T G C T C A T A G C C T T T A C T A A G A A G A G A C C A C T G C T G T T G C A G A G C T C A G A A A C T G A T G   660
QY 1351  G T T C A A G C A G C A G A T C T T C T T T C C A T T C A T A T T C A T T C A T C A T C A T C G C A T C G C A T C C A G G C C   1410
Db      661  G T C C A G G C G C A G A C C T T C T T T C T G C C A T T C A C C T C A T T G C A C C A G C C A T C C A G G C T   720
QY 1411  C 1411
Db      721  C 721

```

```

RESULT 13
CA977646
LOCUS
DEFINITION
  CA977646 900 bp mRNA linear EST 06-JAN-2003
  AGENCOURT 11295472 NIH_MGC_164 Mus musculus cDNA clone
  IMAGE:30146905 5', mRNA sequence.
ACCESSION
  CA977646
VERSION
  CA977646.1 GI:27510300

```

```

KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 900)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe and Dr. Mina
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM0063 row: f column: 02
High quality sequence stop: 655.
FEATURES
source
1..900
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30146905"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 164"
Note="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;
non-normalized full-length enriched library from pooled
mouse embryonic limb, maxilla and mandible, day 10.5 and
11.5 (size selected for the 0.5-1 kb fragments) Cloned
directionally, priming method: Oligo-dT. cDNA enrichment:
>1k bp, Average insert size 1.8k bp. Priming sequence:
5'GACTAGTTCTAGATCGCGAGCGGCCG(T) 3'. Tissue contributed
by, David Rowe. Library constructed by ResGen, Invitrogen
Corp."
ORIGIN
Query Match 19.4%; Score 599.2; DB 14; Length 900;
Best Local Similarity 85.5%; Pred. No. 1.3e-101;
Matches 700; Conservative 0; Mismatches 94; Indels 25; Gaps 2;
QY 1535 ACTATTTTGCAGATTATAGATAGATAAATAAAGTCTGTGAGGTGCTGGAATTTAAACA 1594
Db 3 ACTATTTCTCAGATTATAGACAGATAAATAAAGTCTGTGAGGTGCTGGAATTTAAACA 62
QY 1595 ATTACATGTATCTCGGATTTTCTGTGAATTTAGTGAACAGTCAACCATGTGTTCTTT 1654
Db 63 ACTTACATGTATCTCGGATTTTCTGTGAATTTAGTGAACAGTCAACCATGTGTTCTTT 122
QY 1655 CAAGATCTCTGTTAACAACCATTTCTCGTGGTGGATTAACAATAAGTCTTTGGAATCATC 1714
Db 123 CAAGATCTCTGTTAACAACCATTTCTCGTGGTGGATTAACAATAAGTCTTTGGAATCATC 182
QY 1715 TCATCGACAGATGCTGAAGATGACATCTCGTCTTTTGTGATGATCCCGAGTCTTTC 1774
Db 183 TCATCGACAGATGCTGAAGATGACATCTCGTCTTTTGTGATGATCCCGAGTCTTTC 240
QY 1775 CCCCAGTCTACCTATATAATAATCACCAGGCTAAGGACTGTATCGACTCTTTGTTTAC 1834
Db 241 CCCCAGTCTACCTATATAATAATCACCAGGCTAAGGACTGTATCGACTCTTTGTTTAC 300
QY 1835 TCATGTGTTGGCCATCTCTGTAGTCTTATTCAGATCCTATGATGATCAACAGGCTCGACA 1894
Db 301 TCATGTGTTGGCCGCTTTTGTAGTCTTGTTCAGATCCTATGATGATCAACAGGCTCGACA 360
QY 1895 GAGAGATAGCTTGTGATATTTCTGAGGATTTGCCACCTTGCAGGATGAGTTATGATGAC 1954
Db 361 GAGAGATAGCTTGTGATATTTCTGAGGATTTGCCACCTTGCAGGATGAGTTATGATGAC 410
QY 1955 ATTTTATTTTAAAGCAGAGAGGTTGATGAGCGCTTACACCATGCTGTGTTAAACAG 2014

```

11.5 (size selected for the 0.5-1 kb fragments) Cloned
 directionally, priming method: Oligo-dT. cDNA enrichment:
 >1k bp, Average insert size 1.8k bp. Priming sequence:
 5'GACTAGTTCTAGATCGCGAGCGGCCG(T) 3'. Tissue contributed
 by, David Rowe. Library constructed by ResGen, Invitrogen
 Corp."

Query Match 19.4%; Score 599.2; DB 14; Length 900;
 Best Local Similarity 85.5%; Pred. No. 1.3e-101;
 Matches 700; Conservative 0; Mismatches 94; Indels 25; Gaps 2;

1535 ACTATTTTGCAGATTATAGATAGATAAATAAAGTCTGTGAGGTGCTGGAATTTAAACA 1594
 3 ACTATTTCTCAGATTATAGACAGATAAATAAAGTCTGTGAGGTGCTGGAATTTAAACA 62
 1595 ATTACATGTATCTCGGATTTTCTGTGAATTTAGTGAACAGTCAACCATGTGTTCTTT 1654
 63 ACTTACATGTATCTCGGATTTTCTGTGAATTTAGTGAACAGTCAACCATGTGTTCTTT 122
 1655 CAAGATCTCTGTTAACAACCATTTCTCGTGGTGGATTAACAATAAGTCTTTGGAATCATC 1714
 123 CAAGATCTCTGTTAACAACCATTTCTCGTGGTGGATTAACAATAAGTCTTTGGAATCATC 182
 1715 TCATCGACAGATGCTGAAGATGACATCTCGTCTTTTGTGATGATCCCGAGTCTTTC 1774
 183 TCATCGACAGATGCTGAAGATGACATCTCGTCTTTTGTGATGATCCCGAGTCTTTC 240
 1775 CCCCAGTCTACCTATATAATAATCACCAGGCTAAGGACTGTATCGACTCTTTGTTTAC 1834
 241 CCCCAGTCTACCTATATAATAATCACCAGGCTAAGGACTGTATCGACTCTTTGTTTAC 300
 1835 TCATGTGTTGGCCATCTCTGTAGTCTTATTCAGATCCTATGATGATCAACAGGCTCGACA 1894
 301 TCATGTGTTGGCCGCTTTTGTAGTCTTGTTCAGATCCTATGATGATCAACAGGCTCGACA 360
 1895 GAGAGATAGCTTGTGATATTTCTGAGGATTTGCCACCTTGCAGGATGAGTTATGATGAC 1954
 361 GAGAGATAGCTTGTGATATTTCTGAGGATTTGCCACCTTGCAGGATGAGTTATGATGAC 410
 1955 ATTTTATTTTAAAGCAGAGAGGTTGATGAGCGCTTACACCATGCTGTGTTAAACAG 2014

11.5 (size selected for the 0.5-1 kb fragments) Cloned
 directionally, priming method: Oligo-dT. cDNA enrichment:
 >1k bp, Average insert size 1.8k bp. Priming sequence:
 5'GACTAGTTCTAGATCGCGAGCGGCCG(T) 3'. Tissue contributed
 by, David Rowe. Library constructed by ResGen, Invitrogen
 Corp."

Query Match 19.4%; Score 599.2; DB 14; Length 900;
 Best Local Similarity 85.5%; Pred. No. 1.3e-101;
 Matches 700; Conservative 0; Mismatches 94; Indels 25; Gaps 2;

1535 ACTATTTTGCAGATTATAGATAGATAAATAAAGTCTGTGAGGTGCTGGAATTTAAACA 1594
 3 ACTATTTCTCAGATTATAGACAGATAAATAAAGTCTGTGAGGTGCTGGAATTTAAACA 62
 1595 ATTACATGTATCTCGGATTTTCTGTGAATTTAGTGAACAGTCAACCATGTGTTCTTT 1654
 63 ACTTACATGTATCTCGGATTTTCTGTGAATTTAGTGAACAGTCAACCATGTGTTCTTT 122
 1655 CAAGATCTCTGTTAACAACCATTTCTCGTGGTGGATTAACAATAAGTCTTTGGAATCATC 1714
 123 CAAGATCTCTGTTAACAACCATTTCTCGTGGTGGATTAACAATAAGTCTTTGGAATCATC 182
 1715 TCATCGACAGATGCTGAAGATGACATCTCGTCTTTTGTGATGATCCCGAGTCTTTC 1774
 183 TCATCGACAGATGCTGAAGATGACATCTCGTCTTTTGTGATGATCCCGAGTCTTTC 240
 1775 CCCCAGTCTACCTATATAATAATCACCAGGCTAAGGACTGTATCGACTCTTTGTTTAC 1834
 241 CCCCAGTCTACCTATATAATAATCACCAGGCTAAGGACTGTATCGACTCTTTGTTTAC 300
 1835 TCATGTGTTGGCCATCTCTGTAGTCTTATTCAGATCCTATGATGATCAACAGGCTCGACA 1894
 301 TCATGTGTTGGCCGCTTTTGTAGTCTTGTTCAGATCCTATGATGATCAACAGGCTCGACA 360
 1895 GAGAGATAGCTTGTGATATTTCTGAGGATTTGCCACCTTGCAGGATGAGTTATGATGAC 1954
 361 GAGAGATAGCTTGTGATATTTCTGAGGATTTGCCACCTTGCAGGATGAGTTATGATGAC 410
 1955 ATTTTATTTTAAAGCAGAGAGGTTGATGAGCGCTTACACCATGCTGTGTTAAACAG 2014

Search completed: August 11, 2004, 04:18:40
Job time : 8430.09 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 10, 2004, 17:10:17 ; Search time 1241.45 Seconds
(without alignments)
10594.393 Million cell updates/sec

Title: US-10-001-857-42

Perfect score: 3096

Sequence: 1 ttctcagcaactccagg.....atttcttgacaaaaaaa 3096

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N Geneseq 29Jan04: *
1: geneseqn1980s: *
2: geneseqn1990s: *
3: geneseqn2000s: *
4: geneseqn2001as: *
5: geneseqn2001bs: *
6: geneseqn2002s: *
7: geneseqn2003as: *
8: geneseqn2003bs: *
9: geneseqn2003cs: *
10: geneseqn2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3096	100.0	3096	6	ABQ75303	Abq75303 Human lun
2	2977.2	96.2	3029	9	ADC30041	Adc30041 Human nov
3	2577.8	83.3	2668	5	ABV29823	Abv29823 Human pro
4	2577.8	83.3	2668	5	ABV23940	Abv23940 Human pro
5	2417.2	78.1	2488	6	AAL49929	Aal49929 Human mol
6	1799.2	58.1	2543	9	ADB53632	Adb53632 Primary r
7	717.4	23.2	719	6	ABSS1309	Abss1309 cDNA enco
8	451.2	14.6	482	8	ACH33702	Ach33702 Human end
9	451.2	14.6	482	9	ADC31969	Adc31969 Human nov
10	350.8	11.3	398	7	ABX50281	Abx50281 Bovine ES
11	343.8	11.1	799	6	ABQ51659	Abq51659 Oligonuc
12	343.8	11.1	799	6	ABQ51658	Abq51658 Oligonuc
13	292.4	9.4	356	5	ABV15331	Abv15331 Human pro
14	288.8	9.3	439	5	ABV45182	Abv45182 Human pro
15	288.8	9.3	439	5	ABV36125	Abv36125 Human pro
16	274.2	8.9	799	6	ABQ51661	Abq51661 Oligonuc
17	274.2	8.9	799	6	ABQ51660	Abq51660 Oligonuc
18	230.2	7.4	303	5	ABV06162	Abv06162 Human pro
19	217	7.0	2515	4	ABL07143	AbL07143 Drosophil
20	140.4	4.5	151	6	ABL075302	AbL075302 Human lun
21	90.6	2.9	3401	4	ABL05906	AbL05906 Drosophil
22	90.6	2.9	4847	4	ABL07142	AbL07142 Drosophil
23	87.8	2.8	815	6	ABQ75372	Abq75372 Human lun

24	66.6	2.2	2000	7	ADA71938	Ada71938 Rice gene
25	60	1.9	60	6	ABN43747	Abn43747 Human spl
c	58.6	1.9	600	6	ABQ52497	Abq52497 Oligonuc
26	58.6	1.9	600	6	ABQ52496	Abq52496 Oligonuc
27	58.6	1.9	600	6	ABQ52496	Abq52496 Oligonuc
28	58.2	1.9	3163	9	ADC87060	Adc87060 Human GPC
29	56.4	1.8	320	3	AAA38183	Aaa38183 Primer us
30	56.2	1.8	5452	9	ADC86736	Adc86736 Human GPC
31	56	1.8	451	6	ABT10409	Abt10409 Human bre
32	56	1.8	1117	9	ADC86688	Adc86688 Human GPC
33	56	1.8	1286	6	ABI99656	Abi99656 Mouse isc
c	55.4	1.8	365	5	ABV54856	Abv54856 Human pro
34	55.2	1.8	320	3	AAA38186	Aaa38186 Primer us
35	54.8	1.8	320	3	AAA38185	Aaa38185 Primer us
c	54.8	1.8	615	7	ACA23975	Aca23975 Prokaryot
36	54.6	1.8	840	6	ABQ35494	Abq35494 Oligonuc
37	54.6	1.8	840	6	ABQ35495	Abq35495 Oligonuc
c	54.6	1.8	1416	7	ABZ20967	Abz20967 Animal te
38	54.6	1.8	12733	6	ABK98631	Abk98631 Vector pE
39	54.6	1.8	12733	8	ACDI3882	Acdi3882 L. lactis
40	54.6	1.8	12739	6	ABK98592	Abk98592 Vector pE
41	54.6	1.8	12739	8	ACDI3843	Acdi3843 Plasmid p
42	54.6	1.8	318	3	AAA38184	Aaa38184 Primer us
43	54.6	1.8	318	3	AAA38184	Aaa38184 Primer us
44	54.6	1.8	318	3	AAA38184	Aaa38184 Primer us
45	54.2	1.8	318	3	AAA38184	Aaa38184 Primer us

ALIGNMENTS

RESULT 1	ABQ75303	
ID	ABQ75303 standard; cDNA; 3096 BP.	
XX	ABQ75303;	
AC	ABQ75303;	
XX	05-NOV-2002 (first entry)	
DT	Human lung specific nucleic acid sequence SEQ ID NO:42.	
XX	Human; lung; lung specific nucleic acid; LSNA; lung specific protein; LSP; cytosolic; gene therapy; vaccine; metastasis; lung cancer; squamous cell carcinoma; gene; chromosome 9; ss.	
KW	Homo sapiens.	
XX	WO200264788-A2.	
PN	22-AUG-2002.	
XX	20-NOV-2001; 2001WO-US045080.	
XX	20-NOV-2000; 2000US-0252054P.	
XX	(DIAD-) DIADEXUS INC.	
XX	Macina RA, Recipon H, Chen S, Sun Y, Liu C;	
XX	WPI; 2002-657601/70.	
XX	New lung specific nucleic acid useful in gene therapy or as vaccines for treating lung cancer (e.g. squamous cell carcinoma) or non-cancerous lung diseases, as well as for diagnosing, monitoring or staging these diseases.	
PT	Claim 1; Page 168-169; 282pp; English.	
XX	The present invention describes an isolated lung specific nucleic acid (LSNA) comprising a sequence that: (a) encodes any of the 93 amino acid sequences comprising 17 - 733 amino acids, given in ABP52873 to ABP52965; (b) comprises any of 115 sequences comprising 148 - 3193 base pairs (bp), given in ABQ75262 to ABQ75376; (c) selectively hybridizes to (a) or (b); or (d) has 60 % sequence identity to (a) or (b). LSNA and lung specific protein (LSP) sequences have cytostatic activity and can be used in gene therapy and vaccines. LSNA and LSPs are useful for diagnosing and monitoring the presence and metastases of lung cancer in a patient. An	

QY 1981 TGATGACGGCTTCACACCATGCTGTTGAACAGGAACCCAAAGGCAACATTTGGCCTG 2040
Db 1981 TGATGACGGCTTCACACCATGCTGTTGAACAGGAACCCAAAGGCAACATTTGGCCTG 2040
QY 2041 TTTAGGTACCTGGGTCTCTTACCAATAACCTTCGCATTTATGATACAGTACCTTCTAACTGG 2100
Db 2041 TTTAGGTACCTGGGTCTCTTACCAATAACCTTCGCATTTATGATACAGTACCTTCTAACTGG 2100
QY 2101 CTTTGAATGGAACTCTACAGTATGACAGTACTATTACATATATTGGTATCTCTCTGA 2160
Db 2101 CTTTGAATGGAACTCTACAGTATGACAGTACTATTACATATATTGGTATCTCTCTGA 2160
QY 2161 ATTCTTTTACCATGCTTGTATCAACATTTAGTCGTCGCGATGGCTCTCAATGGCAGA 2220
Db 2161 ATTCTTTTACCATGCTTGTATCAACATTTAGTCGTCGCGATGGCTCTCAATGGCAGA 2220
QY 2221 GGAAGGATAATGGAAGCAGCAGCAAGAAAGCCGTAGTAGTAAAGAAAAACAAAGAAAAAAA 2280
Db 2221 GGAAGGATAATGGAAGCAGCAGCAAGAAAGCCGTAGTAGTAAAGAAAAACAAAGAAAAAAA 2280
QY 2281 GAAAGTTCGCCCATTTAGCCGAGAGATCAAAATGAGCCAAAGCATATCAGAACATGTGTC 2340
Db 2281 GAAAGTTCGCCCATTTAGCCGAGAGATCAAAATGAGCCAAAGCATATCAGAACATGTGTC 2340
QY 2341 TGAATGTTTTAAACCAATGTTAGCATTTGACATGGACGGCAAACTAGTAAACCGAGTT 2400
Db 2341 TGAATGTTTTAAACCAATGTTAGCATTTGACATGGACGGCAAACTAGTAAACCGAGTT 2400
QY 2401 TGAGCTTGATGTAAGCAACAGTTCCGTTATGAACACAGGTTTGCCTCAATTCACACAGTGTGAT 2460
Db 2401 TGAGCTTGATGTAAGCAACAGTTCCGTTATGAACACAGGTTTGCCTCAATTCACACAGTGTGAT 2460
QY 2461 GACCCGCCGCCAGTGCATCTTACAGTTCAAGGAATGTCGACCTCAATAAATATAG 2520
Db 2461 GACCCGCCGCCAGTGCATCTTACAGTTCAAGGAATGTCGACCTCAATAAATATAG 2520
QY 2521 CCCTCTCTCTCAGTCTCTCTCAACTGTATGCGCAGCTTAGTAAAGCACTTTCAACAGGCAAA 2580
Db 2521 CCCTCTCTCTCAGTCTCTCTCAACTGTATGCGCAGCTTAGTAAAGCACTTTCAACAGGCAAA 2580
QY 2581 AATGATATTGGAATAATCTTCTTAACCGGACCATGAGGTTAATAGAAATTTTAAAGTTGC 2640
Db 2581 AATGATATTGGAATAATCTTCTTAACCGGACCATGAGGTTAATAGAAATTTTAAAGTTGC 2640
QY 2641 CAACCCCACTTTCTGTTATGAGTTATTTGGCAGAGGACAAAGAAAGGAACTTAAAGT 2700
Db 2641 CAACCCCACTTTCTGTTATGAGTTATTTGGCAGAGGACAAAGAAAGGAACTTAAAGT 2700
QY 2701 TCCTCTCTGAATTTGATTTCTCTCTCATATAATATTTTCTGTTGTGAAACTCTGTTGAGA 2760
Db 2701 TCCTCTCTGAATTTGATTTCTCTCTCATATAATATTTTCTGTTGTGAAACTCTGTTGAGA 2760
QY 2761 GAGACTGGGAGGTGGCCATAAAGGGGACAGGCTCTTCTTTCAGACCCCACTCTTAGAGGG 2820
Db 2761 GAGACTGGGAGGTGGCCATAAAGGGGACAGGCTCTTCTTTCAGACCCCACTCTTAGAGGG 2820
QY 2821 CACATCACAGGCTCCACATCACGGAAGTGCATGATTTCTTGGTGAACAACTCATTA 2880
Db 2821 CACATCACAGGCTCCACATCACGGAAGTGCATGATTTCTTGGTGAACAACTCATTA 2880
QY 2881 TAAGGAATATCTTTAGTTTGGACAGCTTATATACATGAATGAAACTGCTGTTTAAAG 2940
Db 2881 TAAGGAATATCTTTAGTTTGGACAGCTTATATACATGAATGAAACTGCTGTTTAAAG 2940
QY 2941 TGGTTTATTATGTTCCATGGAAGAAACTGGTCTTATTGTAATGCATTCAGTGAAGCTTATAT 3000
Db 2941 TGGTTTATTATGTTCCATGGAAGAAACTGGTCTTATTGTAATGCATTCAGTGAAGCTTATAT 3000
QY 3001 GGTTTTATTACAGATTAATCAAAATCAATTTTTTATGATGATGAGTGAAGTAAATAGTGT 3060
Db 3001 GGTTTTATTACAGATTAATCAAAATCAATTTTTTATGATGATGAGTGAAGTAAATAGTGT 3060
QY 3061 TTATAAGGTTAATAATTTCTTGACAAAAA 3096

Db 3061 TTATAAGGTTAATAATTTCTTGACAAAAA 3096

RESULT 2

ADC30041
ID ADC30041 standard; cDNA; 3029 BP.

XX AC ADC30041;

XX DT 18-DEC-2003 (first entry)

XX Human novel cDNA sequence, SEQ ID NO:123.

XX Human; diagnostic; drug screening; forensics; gene mapping;
KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;
KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
KW ulcers; osteoporosis; autoimmune disease; cancer;
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;
KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnery;
KW antitumor; osteopathic; immunosuppressive; antiinflammatory; cytostatic;
KW gene therapy; chromosome 9; gene; ss.

XX Homo sapiens.

XX PN WO2003029271-A2.

XX PD 10-APR-2003.

XX PF 24-SEP-2002; 2002WO-US030474.

XX PR 24-SEP-2001; 2001US-0324631P.

XX PA (HYSE-) HYSEQ INC.

PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
PI Haley-Vicente D, Drmanac RT;

XX NPI; 2003-371981/35.

XX P-PSDB; ADC31012.

XX New polynucleotide and polypeptide useful for diagnosing, preventing or
PT treating conditions such as neurodegenerative diseases, anemias, platelet
PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
PT cancer.

XX Claim 1; SEQ ID NO 123; 1185pp; English.

XX The invention relates to 971 novel human cDNA sequences (ADC29919-
CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The
CC invention also relates to nucleic acid sequences over 99% identical with
CC the novel human cDNAs. The invention additionally encompasses expression
CC vectors and host cells comprising a nucleic acid of the invention; the
CC recombinant production of a polypeptide of the invention; an antibody
CC against a polypeptide of the invention; a method of detecting
CC polynucleotides or polypeptides of the invention; and methods of
CC identifying a compound which binds to a polypeptide of the invention. The
CC invention further discloses methods of preventing, treating or
CC ameliorating a medical condition; kits comprising polynucleotide probes
CC and/or monoclonal antibodies for carrying out the methods of the
CC invention; methods for the identification of compounds that modulate the
CC expression or activity of the polynucleotide and/or polypeptide; and 767
CC contig sequences corresponding to the cDNA sequences of the invention
CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
CC -ADC33394). The nucleic acids and polypeptides of the invention are
CC useful in diagnostics, drug screening, forensics, gene mapping, in the
CC identification of mutations responsible for genetic disorders or other
CC traits, for assessing biodiversity, and in producing many other types of
CC data and products dependent on DNA and amino acid sequences. They are
CC also used for treating diseases such as Parkinson's disease, Alzheimer's
CC disease and other neurodegenerative diseases, anaemia, platelet
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or

CC cancer. The nucleic acids may also be used as hybridisation probes or
CC primers, and in the recombinant production of a protein. The polypeptides
CC are also useful in generating antibodies, as molecular weight markers,
CC and as food supplements. The present sequence represents a specifically
CC claimed human cDNA sequence of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 3029 BP; 895 A; 593 C; 739 G; 802 T; 0 U; 0 Other;

Query Match 96.2%; Score 2977.2; DB 9; Length 3029;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 3023; Conservative 0; Mismatches 3; Indels 24; Gaps 2;

Qy	48	TCAGGACAGCTAGCAGCGCTTCTCTCGAGCTTGGCTGCTCGTCTCGCAACC	107
Db	1	TCAGGACAGCTAGCAGCGCTTCTCTCGAGCTTGGCTGCTCGTCTCGCAACC	60
Qy	108	ACTAAGGTCTACGAAACCTCCAGGTTTCTCGCTTCGGCTTCACCTTTCTAAGAA	167
Db	61	ACTAAGGTCTACGAAACCTCCAGGTTTCTCTCGCTTCGGCTTCACCTTTCTAAGAA	120
Qy	168	TTCCACAGAGGCGAGCGAGCGGGCGGGCTCTGAGACTCCGGGCTCCGGCTCTTTCGG	227
Db	121	TTCCACAGAGGCGAGCGAGCGGGCGGGCTCTGAGACTCCGGGCTCCGGCTCTTTCGG	180
Qy	228	GAACGGCCCACTACCCAGGACTCCGACAGAGGGTGAAGAAAGATACTTCCGGTCTCGG	287
Db	181	GAACGGCCCACTACCCAGGACTCCGACAGAGGGTGAAGAAAGATACTTCCGGTCTCGG	240
Qy	288	ATCGTCTCTAATCTCGCAGAGAGAGCGGCGGCCATCGCCGAAACGAGGGCGGTGC	347
Db	241	ATCGTCTCTAATCTCGCAGAGAGAGCGGCGGCCATCGCCGAAACGAGGGCGGTGC	300
Qy	348	GAGGAGGGGGTGTGGCGGGGAGCGGAAAGTCCCGGGAGTAAAGGAGAGGGGCGGGG	407
Db	301	GAGGAGGGGGTGTGGCGGGGAGCGGAAAGTCCCGGGAGTAAAGGAGAGGGGCGGGG	360
Qy	408	TCGGCGTCCCGGGCATACGATGCGTCACTGCTGCGCGGTGGGCTGGGCTGAGAGGGA	467
Db	361	TCGGCGTCCCGGGCATACGATGCGTCACTGCTGCGCGGTGGGCTGGGCTGAGAGGGA	420
Qy	468	GGGGCGGGCGGGCGGAGGCGGGTCTGTTATTTCCGTGTCGACAGTGGTGGCGGC	527
Db	421	GGGGCGGGCGGGCGGAGGCGGGTCTGTTATTTCCGTGTCGACAGTGGTGGCGGC	480
Qy	528	GCGGTGACACCGGAGAGTAGGCATATGTTATGAAAGCTTCTGTAGATGATGACGA	587
Db	481	GCGGTGACACCGGAGAGTAGGCATATGTTATGAAAGCTTCTGTAGATGATGACGA	540
Qy	588	TTGAGGATGGAGCTCAGTATGCGGAAAGAAATGAGAGAAAGCAATACAACTGGGTGA	647
Db	541	TTGAGGATGGAGCTCAGTATGCGGAAAGAAATGAGAGAAAGCAATACAACTGGGTGA	600
Qy	648	CATTACCAAGATTTGAGAGAGCTTGCAGAAATTAAGTTGGGAGAACTACTCATGA	707
Db	601	CATTACCAAGATTTGAGAGAGCTTGCAGAAATTAAGTTGGGAGAACTACTCATGA	660
Qy	708	TAAGCTATTGCTCTTTTGAAGCCATGCTGCTATTGAAATGATGATGCCAAGATGA	767
Db	661	TAAGCTATTGCTCTTTTGAAGCCATGCTGCTATTGAAATGATGATGCCAAGATGA	720
Qy	768	TGCTGGCATGATTGAAACCAAGTTAATCGAAAGTTCTCAATTTTGAACAGCTATCAA	827
Db	721	TGCTGGCATGATTGAAACCAAGTTAATCGAAAGTTCTCAATTTTGAACAGCTATCAA	780
Qy	828	GGATGGCACTATTAATAATTAAGATCTCACTTGGCTGAACCTGATAGGATTTATGATAC	887
Db	781	GGATGGCACTATTAATAATTAAGATCTCACTTGGCTGAACCTGATAGGATTTATGATAC	840
Qy	888	ATGTTTTGCTGTTTGAATTAAGCTGTTGAGAGCCATTCATCTGGCACAGCATGTTTAC	947

Db	841	ATGTTTTGCTGTTTGAATTAACGTGTTAGAGGCCATTCTACTGGCACAGACAGTATTAC	900
Qy	948	GTGCTTTACATTCATTAATCCAGACTTTATAGAGATCCCTGCTATGAGGCTTTTGCCT	1007
Db	901	GTGCTTTACATTCATTAATCCAGACTTTATAGAGATCCCTGCTATGAGGCTTTTGCCT	960
Qy	1008	GGGAATCTTGAATAATCTGTGACATTCGAAGGAAAAAGTAAATAAAGCTGCTGTTTTGA	1067
Db	961	GGGAATCTTGAATAATCTGTGACATTCGAAGGAAAAAGTAAATAAAGCTGCTGTTTTGA	1020
Qy	1068	AGAGAAAGATTTTCAGTCAATGACTTATGAAATTTAAATGGCTAACAGTGTGACAGATCT	1127
Db	1021	AGAGAAAGATTTTCAGTCAATGACTTATGAAATTTAAATGGCTAACAGTGTGACAGATCT	1080
Qy	1128	TCGAGTTACAGGCAATGCTAAAGATCTGGAGGATGACATGCAAGAGAGAGTAAAGAGTAC	1187
Db	1081	TCGAGTTACAGGCAATGCTAAAGATCTGGAGGATGACATGCAAGAGAGAGTAAAGAGTAC	1140
Qy	1188	TCGAAGTCGACAGGAGAAAGAGATCCAGAAAGTTGAACCTAGAAACACCAACAATGTTT	1247
Db	1141	TCGAAGTCGACAGGAGAAAGAGATCCAGAAAGTTGAACCTAGAAACACCAACAATGTTT	1200
Qy	1248	AGCAGTATTCAGCAGAGTGAATAATTAATCTGCTGTTTACTGACAGCTTATAGCCTTTAC	1307
Db	1201	AGCAGTATTCAGCAGAGTGAATAATTAATCTGCTGTTTACTGACAGCTTATAGCCTTTAC	1260
Qy	1308	TAAGAAAGAGACAGTCTGTTGAGAGCTCAAAAATGATGTTCAAGCAGACAGATCT	1367
Db	1261	TAAGAAAGAGACAGTCTGTTGAGAGCTCAAAAATGATGTTCAAGCAGACAGATCT	1320
Qy	1368	TCCTTCTGCCATTCATTAATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT	1427
Db	1321	TCCTTCTGCCATTCATTAATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT	1380
Qy	1428	AGGAGATCAATCAATATGATGGGTTTGAACCCCTTGTGAAACAGAGGCTACTTCCACC	1487
Db	1381	AGGAGATCAATCAATATGATGGGTTTGAACCCCTTGTGAAACAGAGGCTACTTCCACC	1440
Qy	1488	TACCTTCCCTCGATATGCAAAATAATTAAGAGGAGAAATGTTGAATTTTTCGAAG	1547
Db	1441	TACCTTCCCTCGATATGCAAAATAATTAAGAGGAGAAATGTTGAATTTTTCGAAG	1500
Qy	1548	ATTAATAGATAGATAAATACTGCTGTGAGGTTTGAATTTAAACAAATTTACATTTGAT	1607
Db	1501	ATTAATAGATAGATAAATACTGCTGTGAGGTTTGAATTTAAACAAATTTACATTTGAT	1560
Qy	1608	CCTGGAATTTTCTGTGAATTTAGTGAACAGTCAACATGTTCTTTTCAAGATCTCTGTT	1667
Db	1561	CCTGGAATTTTCTGTGAATTTAGTGAACAGTCAACATGTTCTTTTCAAGATCTCTGTT	1620
Qy	1668	ACAAACCACTTTCCTGGTGGATACAAAGAGTCTTTTGGAACTCATCTCATGCAAGACAT	1727
Db	1621	ACAAACCACTTTCCTGGTGGATACAAAGAGTCTTTTGGAACTCATCTCATGCAAGACAT	1680
Qy	1728	GCTGAAAGATGACATTCG-CTCTTTTGTGAGATCCTCCGAGTGGTCTTCCCCAAGTGCTA	1786
Db	1681	GCTGAAAGATGACATTCGCGTCTTTTGTGAGATCCTCCGAGTGGTCTTCCCCAAGTGCTA	1740
Qy	1787	CCTATATTAATAATCAACAGGCTAGGACTGTATGACTCTCTTTTGTACTCTGTTTGG	1846
Db	1741	CCTATATTAATAATCAACAGGCTAGGACTGTATGACTCTCTTTTGTACTCTGTTTGG	1800
Qy	1847	GCCATTCTGTAGTCTTATTCAGATCCATGGACATAAAGGGCTCCAGAGAGAGATAAGCT	1906
Db	1801	GCCATTCTGTAGTCTTATTCAGATCCATGGACATAAAGGGCTCCAGAGAGAGATAAGCT	1860
Qy	1907	TGCTCATATTTCTGAGGAAATTTGCCACCTTCAGAGATGAGTTTATGACATTTATTTAA	1966
Db	1861	TGCTCATATTTCTGAGGAAATTTGCCACCTTCAGAGATGAGTTTATGACATTTATTTAA	1898
Qy	1967	TAGCAGAGAGAGTTGATGAGCGGTTTCAACACCTGCTTGAACAGAGAGAGAGAGG	2026
Db	1899	-AGCAGAGAGAGTTGATGAGCGGCTTCACACCTGCTTGAACAGAGAGAGAGAGG	1957

Db 2488 TATCCGAGAAAATGAGAAAGCAATCAAACTGGGTGACATTCACCAAGATTTTGA 2429
Qy 666 AGAAGCTTGCGAGAAATAAAGTTGGGAGAACTACTCTCATGATAAGCTATTGGTCTTTT 725
Db 2428 AGAAGCTTGCGAGAAATAAAGTTGGGAGAACTACTCTCATGATAAGCTATTGGTCTTTT 2369
Qy 726 TGAAGCCATGCTCTGCTATTGAAATGATGGATCCCAAGATGATGCTGGCATGATTGGAAA 785
Db 2368 TGAAGCCATGCTCTGCTATTGAAATGATGGATCCCAAGATGATGCTGGCATGATTGGAAA 2309
Qy 786 CCAAGTTAAATCGAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATTAAAT 845
Db 2308 CCAAGTTAAATCGAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATTAAAT 2249
Qy 846 TAAAGATCTCACTTGGCTGAACTGATAGGGATTTAGGATACATGTTTTGCTGCTTTGAT 905
Db 2248 TAAAGATCTCACTTGGCTGAACTGATAGGGATTTAGGATACATGTTTTGCTGCTTTGAT 2189
Qy 906 AACGTGTTAGAGGCCATTCACTGGCACACAGAGTATTTACGTGCCCTTTACATTCATA 965
Db 2188 AACGTGTTAGAGGCCATTCACTGGCACACAGAGTATTTACGTGCCCTTTACATTCATA 2129
Qy 966 TCCAGACTTTATAGAAGATCCTGCTATGAAGGCTTTTGTCTGGGAATCTTGAATACTG 1025
Db 2128 TCCAGACTTTATAGAAGATCCTGCTATGAAGGCTTTTGTCTGGGAATCTTGAATACTG 2069
Qy 1026 TGACATTTGAAGGAAAGTAAATAAGCTGCTGTTTTTGAAGAGAAATTTTCAGTC 1085
Db 2068 TGACATTTGAAGGAAAGTAAATAAGCTGCTGTTTTTGAAGAGAAATTTTCAGTC 2009
Qy 1086 AATGACTTATGGAATTTAAATGGCTAACAGTGTGACAGATCTTCGAGTTACAGCATGCT 1145
Db 2008 AATGACTTATGGAATTTAAATGGCTAACAGTGTGACAGATCTTCGAGTTACAGCATGCT 1949
Qy 1146 AAAAGATGTGGAGATGACATGCAAAAGAGTAGTAAGAGTACTCCAAAGTCGACAAGGAGA 1205
Db 1948 AAAAGATGTGGAGATGACATGCAAAAGAGTAGTAAGAGTACTCCAAAGTCGACAAGGAGA 1889
Qy 1206 AGAAGAGATCCAGAACTGGAATAGAACACCAACAAATGTTTACAGATATTCAGCAGAT 1265
Db 1888 AGAAGAGATCCAGAACTGGAATAGAACACCAACAAATGTTTACAGATATTCAGCAGAT 1829
Qy 1266 GAAATTTACTGCTGCTGTTACTGACAGTGTATAGCCCTTACTAAGAAAGAGACCACTGC 1325
Db 1828 GAAATTTACTGCTGCTGTTACTGACAGTGTATAGCCCTTACTAAGAAAGAGACCACTGC 1769
Qy 1326 TGTTCAGAAAGCTCAAAAATTTGATGGTTCAAGCAGCAGATCTTCTTCTGCAATTCATA 1385
Db 1768 TGTTCAGAAAGCTCAAAAATTTGATGGTTCAAGCAGCAGATCTTCTTCTGCAATTCATA 1709
Qy 1386 TTCAATTCATCATGGAATCCAGGCCAGAAATGATACATAAAGGAGATCATCCAATAT 1445
Db 1708 TTCAATTCATCATGGAATCCAGGCCAGAAATGATACATAAAGGAGATCATCCAATAT 1649
Qy 1446 GATGGGTTTTGAACCCCTTGTGAACACAGAGCTACTTCCACCTACCTTCCCTCGATATGC 1505
Db 1648 GATGGGTTTTGAACCCCTTGTGAACACAGAGCTACTTCCACCTACCTTCCCTCGATATGC 1589
Qy 1506 AAAAATAATTAAGGGAAGAAATGGTGAACATAATTTTGAAGATTAATAGATAGAAATAA 1565
Db 1588 AAAAATAATTAAGGGAAGAAATGGTGAACATAATTTTGAAGATTAATAGATAGAAATAA 1529
Qy 1566 AACTGCTGTCAGGTTGTAATTTAAACAATTTACATGTTATCCTGGAATTTTCTGTGA 1625
Db 1528 AACTGCTGTCAGGTTGTAATTTAAACAATTTTACATGTTATCCTGGAATTTTCTGTGA 1469
Qy 1626 ATTTAGTGAACAGTCACCATGCTGTTTTCAAGATCTCTGTACAAACCACTTCCCTGGT 1685
Db 1468 ATTTAGTGAACAGTCACCATGCTGTTTTCAAGATCTCTGTACAAACCACTTCCCTGGT 1409
Qy 1686 GGTAAACAAAAGGCTTTTGAACCTCATCTCATGCAAGAATGGTGAAGATGCATTCG 1745

Db 1408 GGATAACAAAAGGTCTTTTGGAACTCATCTCATGCAAGACATGCTGAAAGATGCACCTCG 1349
Qy 1746 GTCTTTTGTGAGATCCCTCCGAGTCTTTTCCCCCAAGTCTACCTATATATAAATCAACAG 1805
Db 1348 GTCTTTTGTGAG--TCCTCCGGTGTCTTTTCCCCCAAGTCTACCTATATATAAATCAACAG 1291
Qy 1806 GCTTAAGACATGATCGACTCTTTTGTGTACTCAGTGTGTTCGGCCATCTCTGTAGTCTTATT 1865
Db 1290 GCTTAAGACATGATCGACTCTTTTGTGTACTCAGTGTGTTCGGCCATCTCTGTAGTCTTATT 1231
Qy 1866 CAGATCCATCGACATAACAGGGCTCGACAGAGAGATAAGCTTGCTCATATTTCTTTGAGAA 1925
Db 1230 CAGATCCATCGACATAACAGGGCTCGACAGAGAGATAAGCTTGCTCATATTTCTTTGAGAA 1171
Qy 1926 TTTGCCACCTTGGCAGGATGATTTATGACATTTTATTTTAAATAGCAGAGAAAGTTGATG 1985
Db 1170 TTTGCCACCTTGGCAGGATG-----AGGCAGAGAAAGTTGATG 1134
Qy 1986 CAGCGCTTCACCATCTGTTGAAACAGGAAACCCCAAGGCAACATTTGGCCTGTTTAG 2045
Db 1133 CAGCGCTTCACCATCTGTTGAAACAGGAAACCCCAAGGCAACATTTGGCCTGTTTAG 1074
Qy 2046 GTACCTGGGTCTTTTACATAAACCTTCGCATTTATGATACAGTACCTTTCTAAGTGGCTTTG 2105
Db 1073 GTACCTGGGTCTTTTACATAAACCTTCGCATTTATGATACAGTACCTTTCTAAGTGGCTTTG 1014
Qy 2106 AATTGGAACTCTACAGTATGACAGAG--TACTATTATCATATATTGGTATCTCTCTGAATTC 2164
Db 1013 AATTGGAACTCTACAGTATGACAGAGATCTATTACATATATTGGTATCTCTCTGAATTC 954
Qy 2165 CTTTACGCATGGTGTGATGCTCAACATTTAGTCTGCGCATGGCTCTCAATGGCAGAGAA 2224
Db 953 CTTTACGCATGGTGTGATGCTCAACATTTAGTCTGCGCATGGCTCTCAATGGCAGAGAA 894
Qy 2225 AGGATATGAAAGAGCAGCAGAAAGGCGCTAGTAGTAAAAAACAAGAAAAAAGAAAA 2284
Db 893 AGGATATGAAAGAGCAGCAGAAAGGCGCTAGTAGTAAAAAACAAGAAAAAAGAAAA 834
Qy 2285 GTTCGCCCATGAGCCGAGAGATCAATGAGCCCAAGCATATCAGAACATGTGTGCTGGA 2344
Db 833 GTTCGCCCATGAGCCGAGAGATCAATGAGCCCAAGCATATCAGAACATGTGTGCTGGA 774
Qy 2345 ATGTTTTAAACCATGTGTAGCATTTGACATGAGCGCAAGTACGTAAACCGAGTTTGTAG 2404
Db 773 ATGTTTTAAACCATGTGTAGCATTTGACATGAGCGCAAGTACGTAAACCGAGTTTGTAG 714
Qy 2405 CTTGTATAGTGAACAAAGTTTCGGTATGAACACAGGTTTGTCTCCATTCACACAGTGTGATGACC 2464
Db 713 CTTGTATAGTGAACAAAGTTTCGGTATGAACACAGGTTTGTCTCCATTCACACAGTGTGATGACC 654
Qy 2465 CCGCGCCAGTGCACCTACTTTACAGTTCAAGGAATGTCTGACCTCAATTAATATAGCCCT 2524
Db 653 CCGCGCCAGTGCACCTACTTTACAGTTCAAGGAATGTCTGACCTCAATTAATATAGCCCT 594
Qy 2525 CCTCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACCTTTCAACAGGCAAAAAATG 2584
Db 593 CCTCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACCTTTCAACAGGCAAAAAATG 534
Qy 2585 ATATTGAAAATATTCTTAAACCCGACCATGAGGTTAATAGAAATTTTAAAGTTTGCACAA 2644
Db 533 ATATTGAAAATATTCTTAAACCCGACCATGAGGTTAATAGAAATTTTAAAGTTTGCACAA 474
Qy 2645 CCCAACTTTTGTGTTATGAAGTTATTGGCAGGAGGACACAAAAGGATCTTAAAGTTCTCT 2704
Db 473 CCCAACTTTTGTGTTATGAAGTTATTGGCAGGAGGACACAAAAGGATCTTAAAGTTCTCT 414
Qy 2705 CCTGAAATTTGATTTCTCTGCTCATAAATATTTTCTGTGTGAAAATCTGTTTTCAGAGAGA 2764
Db 413 CCTGAAATTTGATTTCTCTGCTCATAAATATTTTCTGTGTGAAAATCTGTTTTCAGAGAGA 354
Qy 2765 CTGGGAGGTGGCCATAAAGGGCAGAGTCTTTTCAGACCCAACTCTTTAGAGGGCACA 2824
Db 353 CTGGGAGGTGGCCATAAAGGGCAGAGTCTTTTCAGACCCAACTCTTTAGAGGGCACA 294

Db 1708 TTCAATTCATATGCGATCCAGGCCAGAGATGATACATAAAGAGATCATCAATAT 1649
QY 1446 GATGGGTTTGAACCCCTTGTGAACCGAGAGTACTTCCACCTACCTTCCCTCGATATGC 1505
Db 1648 GATGGGTTTGAACCCCTTGTGAACCGAGAGTACTTCCACCTACCTTCCCTCGATATGC 1589
QY 1506 AAAAATAATTAAGAGGAAGAAATGGTGAATATTTTTCGAGATTAATAGATAGATAAA 1565
Db 1588 AAAAATAATTAAGAGGAAGAAATGGTGAATATTTTTCGAGATTAATAGATAGATAAA 1529
QY 1566 AACTGTCTGTGAGTGTGTAATTTAAACAAATTTTACATTTGTATCTCGTGA 1625
Db 1528 AACTGTCTGTGAGTGTGTAATTTAAACAAATTTTACATTTGTATCTCGTGA 1469
QY 1526 ATTATGTAACAGTACATGTGTCTTTTCAAGATCTCTGTGAACAAACCTTTCTGTGT 1685
Db 1468 ATTATGTAACAGTACATGTGTCTTTTCAAGATCTCTGTGAACAAACCTTTCTGTGT 1409
QY 1686 GGATAACAAAAGGTCTTTTGAACCTCATCTCAAGAGATGTGTGAAGATGCACCTTCG 1745
Db 1408 GGATAACAAAAGGTCTTTTGAACCTCATCTCAAGAGATGTGTGAAGATGCACCTTCG 1349
QY 1746 GTCTTTTGTGAGTCTTCCGAGTGTCTTTTCCGAGTGTCTACCTATATATAATCACCAG 1805
Db 1348 GTCTTTTGTGAGTCTTCCGAGTGTCTTTTCCGAGTGTCTACCTATATATAATCACCAG 1291
QY 1806 GCTAAGACTGTATCGACTCTTTTGTACTCAGTGTGTTCGGCACTCTGTAGTCTTATT 1865
Db 1290 GCTAAGACTGTATCGACTCTTTTGTACTCAGTGTGTTCGGCACTCTGTAGTCTTATT 1231
QY 1866 CAGATCCATGACATACAGGGTTCGACAGAGATAGAGTGTGTCTATATCTTGAAGAA 1925
Db 1230 CAGATCCATGACATACAGGGTTCGACAGAGATAGAGTGTGTCTATATCTTGAAGAA 1171
QY 1926 TTTGCCACCTTGCAGGTATGATTTATGACATTTTATTTTATAGGCGAGAGAGTGTATG 1985
Db 1170 TTTGCCACCTTGCAGGTATGATTTATGACATTTTATTTTATAGGCGAGAGAGTGTATG 1134
QY 1986 CAGGCTTTCACACCATGCTGTGTGAACAGAGAACCCCAAGCAACATTTGSCCTGTTTAT 2045
Db 1133 CAGGCTTTCACACCATGCTGTGTGAACAGAGAACCCCAAGCAACATTTGSCCTGTTTAT 1074
QY 2046 GTACTGGTCTCTTACATTAACCTTCGATTTATGATACAGTACTCTTAAGTGGCTTTG 2105
Db 1073 GTACTGGTCTCTTACATTAACCTTCGATTTATGATACAGTACTCTTAAGTGGCTTTG 1014
QY 2106 AATTGGAACCTACAGTATGACAGG--TACTATTACATATATTTGGTATCTCTCTGAATTC 2164
Db 1013 AATTGGAACCTACAGTATGACAGGATACTATTACATATATTTGGTATCTCTCTGAATTC 954
QY 2165 CTTTACGCATGTTGATCTCAACATTTGAGTGTGCGGATGCTCTCAATGGCGAGAGAA 2224
Db 953 CTTTACGCATGTTGATCTCAACATTTGAGTGTGCGGATGCTCTCAATGGCGAGAGAA 894
QY 2225 AGGATAATGAAGAGCAGAGAAAGCGGTAGTAGTAAATAAACAAGAAAAAAGAA 2284
Db 893 AGGATAATGAAGAGCAGAGAAAGCGGTAGTAGTAAATAAACAAGAAAAAAGAA 834
QY 2285 GTTCGCCCTTGAAGCGAGAGATCAATAGAGCCAGCATATCAGACATGTGTGTGGA 2344
Db 833 GTTCGCCCTTGAAGCGAGAGATCAATAGAGCCAGCATATCAGACATGTGTGTGGA 774
QY 2345 ATGTTTAAACCATGTGTGATTTGACATGAGCGGCAAGTACGTAAACCGAGTTTGTAG 2404
Db 773 ATGTTTAAACCATGTGTGATTTGACATGAGCGGCAAGTACGTAAACCGAGTTTGTAG 714
QY 2405 CTTGATAGTGAACAGTTCGGTATGAACACAGAGTTTGTCTCCATTCACAGTGTGATGACC 2464
Db 713 CTTGATAGTGAACAGTTCGGTATGAACACAGAGTTTGTCTCCATTCACAGTGTGATGACC 654
QY 2465 CCGCGGCCAGTGCACCTACTTACAGTTCAAGGAATGTCTGACCTCAATATAATATAGCCCT 2524

Db 653 CCGCGCCAGTGCACCTACTTTACAGTTCAAGGAATGTCTGACCTCAATAAATATAGCCCT 594
QY 2525 CTTCTCAGTCTCTCTGAACTGTATGTGGCAGCTAGTAAAGCACCTTTCAACAGGCAAAAATG 2584
Db 593 CTTCTCAGTCTCTCTGAACTGTATGTGGCAGCTAGTAAAGCACCTTTCAACAGGCAAAAATG 534
QY 2585 ATATTGAAAAATATTCTTAACCCGAGCATTGAGGTAAATAGAAATTTTAAAGGTTGCCAAA 2644
Db 533 ATATTGAAAAATATTCTTAACCCGAGCATTGAGGTAAATAGAAATTTTAAAGGTTGCCAAA 474
QY 2645 CCCAACTTTTGTGGTTATGAAGTTTATTTGSCAGAGGACACAAAAAGGAATCTTAAAGTTTCT 2704
Db 473 CCCAACTTTTGTGGTTATGAAGTTTATTTGSCAGAGGACACAAAAAGGAATCTTAAAGTTTCT 414
QY 2705 CCTGAATTTGATTTCTCTGCTCATATAATATTTTCTCTGTTGTAACCTTTTGTGAGAGA 2764
Db 413 CCTGAATTTGATTTCTCTGCTCATATAATATTTTCTCTGTTGTAACCTTTTGTGAGAGA 354
QY 2765 CTGGGAGGTGGCCATAAAGGGGAGAGTCTTTCTTCAGACCCCACTCTTTAGAGGGCACA 2824
Db 353 CTGGGAGGTGGCCATAAAGGGGAGAGTCTTTCTTCAGACCCCACTCTTTAGAGGGCACA 294
QY 2825 TCACAGGCTCCACATCACGGGAAGTGAATGATTTCTTTGGGTAACTCACTTAAAGTGGT 2884
Db 293 TCACAGGCTCCACATCACGGGAAGTGAATGATTTCTTTGGGTAACTCACTTAAAGTGGT 234
QY 2885 GAATACCTTTTGTGGTGTGAGCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 2944
Db 233 GAATACCTTTTGTGGTGTGAGCTTATGATGATGATGATGATGATGATGATGATGATGAT 174
QY 2945 TTATTATGTTCCATGGAAGAACTGGTCTTTTATGATGATGATGATGATGATGATGATGATGAT 3004
Db 173 TTATTATGTTCCATGGAAGAACTGGTCTTTTATGATGATGATGATGATGATGATGATGATGAT 114
QY 3005 TTATTACAGATTTAATCACAATCATTTTATGATGATGATGATGATGATGATGATGATGATGAT 3064
Db 113 TTATTACAGATTTAATCACAATCATTTTATGATGATGATGATGATGATGATGATGATGATGAT 54
QY 3065 AAAGGTTAAATAAATTTCTTCACAAAAA 3095
Db 53 AAAGGTTAAATAAATTTCTTGACAAAAA 23

RESULT 5
AAL49929
ID AAL49929 standard; cDNA; 2488 BP.
XX AAL49929;
AC AC
XX XX
DT 10-DEC-2002 (first entry)
XX
DE Human molecule for disease detection and treatment coding sequence #3.
XX
KW Human; molecule for disease detection and treatment; MDDT; gene therapy;
KW cytosolic; antiarteriosclerotic; hepatotropic; anti-HIV; antiallergic;
KW antiinflammatory; antilasthmatic; cerebroprotective; nootropic;
KW neuroprotective; antiparkinsonian; cardiac; antidiabetic; gene; ss.
OS Homo sapiens.
XX
PN WO200270709-A2.
XX
PD 12-SEP-2002.
XX
PF 08-FEB-2002; 2002WO-US003709.
XX
PR 09-FEB-2001; 2001US-0268117P.
PR 15-FEB-2001; 2001US-0269618P.
PR 23-FEB-2001; 2001US-0271118P.
PR 07-MAR-2001; 2001US-0274486P.
PR 09-MAR-2001; 2001US-0274436P.
PR 28-NOV-2001; 2001US-0334229P.
PR 01-FEB-2002; 2002US-0353284P.


```
Db 1596 TAACTGGCTTTGAATTGGAACCTACAGTATGACGAGTACTATTACATATATTGGTATC 1655
Qy 2154 TCTCTGAATTCCTTTAGCATGGTTGATGTCAACATTGAGTCGTGCGATGCTCTCAA 2213
Db 1656 TCTCTGAATTCCTTTAGCATGGTTGATGTCAACATTGAGTCGTGCGATGCTCTCAA 1715
Qy 2214 TGGCAGAGAAAGATATGAAGCAGCAGCAGAGAGCGGTAGTAGTAAAAAACAAGA 2273
Db 1716 TGGCAGAGAAAGATATGAAGCAGCAGCAGAGAGCGGTAGTAGTAAAAAACAAGA 1775
Qy 2274 AAAAAAAGAAAGTTGCGCCATTGAGCCGAGAGATCACAAATGAGCCAAAGCATATCAGAACA 2333
Db 1776 AAAAAAAGAAAGTTGCGCCATTGAGCCGAGAGATCACAAATGAGCCAAAGCATATCAGAACA 1835
Qy 2334 TGTGTGCTGGAATGTTTAAACCATGTGATGATTTGACATGAGCGGCAAGTAGTAAAC 2393
Db 1836 TGTGTGCTGGAATGTTTAAACCATGTGATGATTTGACATGAGCGGCAAGTAGTAAAC 1895
Qy 2394 CGAAGTTTGAAGTTGATAGTGAACAAGTTGCGTATGAACAGAGTTTCTCCATTCAACA 2453
Db 1896 CGAAGTTTGAAGTTGATAGTGAACAAGTTGCGTATGAACAGAGTTTCTCCATTCAACA 1955
Qy 2454 GTGTGATGACCCCGCCAGTCAGTCACTATTACAGTTCAAGGAAATGTCACCTCAATA 2513
Db 1956 GTGTGATGACCCCGCCAGTCAGTCACTATTACAGTTCAAGGAAATGTCACCTCAATA 2015
Qy 2514 AATATAGCCCTCCTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACTTCAAC 2573
Db 2016 AATATAGCCCTCCTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACTTCAAC 2075
Qy 2574 AGGCAAAAATGATATTGAAAATATTCCTAACCCGACCATGAGGTTAATAGAAATTTAA 2633
Db 2076 AGGCAAAAATGATATTGAAAATATTCCTAACCCGACCATGAGGTTAATAGAAATTTAA 2135
Qy 2634 AGTTTGCAAAACCAACTTTGTGTTATGAAGTTATTGGCAGGAGGACACAAAAGGAAT 2693
Db 2136 AGTTTGCAAAACCAACTTTGTGTTATGAAGTTATTGGCAGGAGGACACAAAAGGAAT 2195
Qy 2694 CTAAAGTTCCCTCGAATTTGATTTCTGCTCATTAATATTTCCTGTTGTAACCTTG 2753
Db 2196 CTAAAGTTCCCTCGAATTTGATTTCTGCTCATTAATATTTCCTGTTGTAACCTTG 2255
Qy 2754 TTTGAGAGAGACTGGGAGGTGGCCATAAAGGGGAGAGTCTTTTCAGACCCCAACTCT 2813
Db 2256 TTTGAGAGAGACTGGGAGGTGGCCATAAAGGGGAGAGTCTTTTCAGACCCCAACTCT 2315
Qy 2814 TAGAGGGCACATCCAGGCTCCACATCACCGGAAAGTGAGATGGAATTTCTGGGTAAACA 2873
Db 2316 TAGAGGGCACATCCAGGCTCCACATCACCGGAAAGTGAGATGGAATTTCTGGGTAAACA 2375
Qy 2874 CTCATTATAGGAATACCTTTTAGTTTGAAGCTTATATGACGCTTATATGACAAATGCTGT 2933
Db 2376 CTCATTATAGGAATACCTTTTAGTTTGAAGCTTATATGACGCTTATATGACAAATGCTGT 2435
Qy 2934 TTTAAAGTGPTTATTATGTTCCATGGGAAGAACTGGTCTTTATTGAAT 2981
Db 2436 TTTAAAGTGPTTATTATGTTCCATGTAAGACACTGGGTTCCATTAAT 2483
```

RESULT 6

```
ADBS53632
ID ADBS53632 standard; DNA; 2543 BP.
XX
AC ADBS53632;
XX
DT 04-DEC-2003 (first entry)
XX
DE Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:4174.
XX
KW toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;
KW toxicity marker; toxicity progression; drug screening;
KW primary rat hepatocyte toxicity modelling; gene; ds.
XX
```

```
OS Rattus norvegicus.
XX
FN WO2003065993-A2.
XX
PD 14-AUG-2003.
XX
PF 04-FEB-2003; 2003WO-US003482.
XX
PR 04-FEB-2002; 2002US-0353171P.
PR 13-MAR-2002; 2002US-0363534P.
PR 08-APR-2002; 2002US-0370248P.
PR 10-APR-2002; 2002US-0371134P.
PR 10-APR-2002; 2002US-0371135P.
PR 10-APR-2002; 2002US-0371150P.
PR 11-APR-2002; 2002US-0371413P.
PR 19-APR-2002; 2002US-0373601P.
PR 19-APR-2002; 2002US-0373602P.
PR 22-APR-2002; 2002US-0374139P.
PR 08-MAY-2002; 2002US-0378370P.
PR 09-MAY-2002; 2002US-0378652P.
PR 09-MAY-2002; 2002US-0378653P.
PR 09-MAY-2002; 2002US-0378665P.
PR 09-JUL-2002; 2002US-0394230P.
PR 09-JUL-2002; 2002US-0394253P.
PR 04-SEP-2002; 2002US-0407688P.
PR 28-JAN-2003; 2003US-0442900P.
XX
(GENE-) GENE LOGIC INC.
XX
FA Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;
PI Elashoff M;
XX
WPI; 2003-731472/69.
```

Determining if a compound induces a toxic effect on a tissue or cell, for identifying hepatotoxic compounds, comprises comparing a gene expression profile of a tissue or cell sample to a database of Tox mean and non-Tox mean values.

Claim 44; SEQ ID NO 4174; 874pp; English.

The present invention describes a method for determining whether a compound induces a toxic effect on a tissue or cell. The method comprises preparing a gene expression profile of a tissue or cell sample exposed to the compound, and comparing the gene expression profile to a database comprising data or information on the Tox mean and non-Tox mean value. The method is useful for predicting or identifying at least one toxic effect, particularly hepatotoxicity, of a test or unknown compound. The genes listed in the specification are useful as diagnostic or toxicity markers for the prediction or identification of the physiological state of tissue or cell sample that has been exposed to a compound, or to identify or predict the toxic effects of a compound or an agent. These may also be used as markers for monitoring toxicity progression or for drug screening. The present sequence represents a primary rat hepatocyte toxicity modelling related gene sequence from the present invention.

Sequence 2543 BP; 725 A; 554 C; 627 G; 637 T; 0 U; 0 Other;

Query Match 58.1%; Score 1799.2; DB 9; Length 2543;
Best Local Similarity 84.0%; Pred. No. 0;
Matches 2139; Conservative 0; Mismatches 343; Indels 64; Gaps 7;

```
Qy 550 GGCATAATGGTTATGAAGCTTCGTAGATCATCAGCATTCAGATGGAGCTCAGTATG 609
Db 58 GGCATAATGGTTATGAAGCCGCGAGTAGATGATCGCTTCGGATGGAGCTCAATGTC 117
Qy 610 CCAGAAAAATGGAGAAAAAGCAATACAACTGGGTGGACATACCCAGATTTTGAAGAA 669
Db 118 CCTCAGAAAAATGGAGAAAAAGTAGCAAGCTGGGTGGACATAACCCAGGACTTTGAAGAT 177
Qy 670 GCTTGTGAGAAATTAAGTTGGGAGAACTACTTTCATGATAGCTATTGGTCTTTTGA 729
Db 178 GCTTGTGAGAGCTGAAGTTGGGAGAACTGCTTCACGATAAGCTGTTGGTCTGTTGAA 237
```


QY 730 GCCATGTCCTGCTATTGAATGATCGATCCCAAGATGATGCTGGCATGATTGGAAACCAA 789
Db 238 GCCATGTCCTGCTATTGAATGATCGATCCCAAGATGATGCTGGCATGATTGGAAACCAA 297
QY 790 GTTAATCGAAAGTCTCAATTTTGAACAAGCTATCAAGGATGCACTATTATTAATTAATA 849
Db 298 GTGNATAGAAAGTCTCAATTTTGAACAAGCCGTCAAGGATGGCACTATTATTAATTAATA 357
QY 850 GATCTCACTTGCCTGAAGTATAGGATATAGGATATACATGTTTTTCTGCTGTTTATTAACG 909
Db 358 GACCTCAGCTTGCCTGAAGTATAGGATATAGGATATACATGTTTTTCTGCTGTTTATTAACG 417
QY 910 TGGTTAAGGCCATTCACCTGGCAGACACAGATATTTAGTGCCTTTACATTCATTAATCCA 969
Db 418 TGGCTGAAGGCCATTCCTTTGGCAGACACAGGATATTTATCATGCTTTATCATTAATCCA 477
QY 970 GACTTTATAGAAGATCCTGCTATGAAGCTTTTCTCTGGGAATCTTGAAGATCTGTGAC 1029
Db 478 GACTTCATAGAGACCTCGCATGAAGCTTTTCTCTGGGAATCTTGAAGATCTGTGAC 537
QY 1030 ATTGCAAGGAAAGTAATAAAGCTGCTGTTTTTGAAGAGGAAGATTTTCAGTCAATG 1089
Db 538 ATTGCAAGGAAAGTAATAAAGCTGCTGTTTTTGAAGAGGAAGATTTTCAGTCAATG 597
QY 1090 ACTTATCGATTTAAATGGCTTAACAGTGTGACAGATCTTCGAGTTACAGGATGCTAAAA 1149
Db 598 ACATATGATTTAAATGGCCCAAGGTGTGACAGATCTTCGAGTTACAGGATGCTCAAA 657
QY 1150 GATGTGAGGATGATCATGCAAGAAAGTAAGAGTACTCGAAGTCGACAGGAAGAA 1209
Db 658 GATGTGAGGATGATCATGCAAGAGAGTAAGAGTACTCGAAGTCGACAGGAAGAG 717
QY 1210 AGAGATCCAGAGTTGAATAGAACACCAATGTTTACAGPATTCAGCAGAGTGAAA 1269
Db 718 AGAGATCCAGAGTTGAATAGAACACCAAGTGTGCGAGCTTGGCAGCGTTTCAGCAGAGTGAAG 777
QY 1270 TTTTACTCGTGTGTACTGACAGTGTCTTATAGCTTTTACTAAGAAAGAGACAGTGTGTT 1329
Db 778 TTTTACCCGAGTACTGTCTGACAGTGTCTATAGCTTTTCAAGAAAGAGACAGTGTGTT 837
QY 1330 GCAGAAAGCTCAAAATATGATGTTTCAAGCAGCAGATCTTCTTCTGCAATTCATTAATCA 1389
Db 838 GCAGAGGCTCAGAAATGATGTTTCCAGCGCAGACCTTCTTCTGCAATTCACCTCA 897
QY 1390 TTGCATCATGGATCCAGGCCAGAAATGATGATCTACAAAAGAGATCATCCAAATATGATG 1449
Db 898 TTGCATCCAGCGCATCCAGGCTCAGAAATGGCACTACCAAGAGAGACCAATCCAAATATGATG 957
QY 1450 GGTTTTGAACCCCTTGTGACACAGAGGCTACTTCCACCTACCTTCCCTCGATATGCAAAA 1509
Db 958 GGTTTTGAACCCCTTGTGACACAGAGATTTACTTCCACCCACCTTCCCTCGATATGCAAAA 1017
QY 1510 ATAAATTAAGAGGAAGAAATGTTGAACATAATTTTGAAGATTAATAGATAGATAAATAAAT 1569
Db 1018 ATAAATTAAGAGGAAGAAATGTTGAACATAATTTTGAAGATTAATAGATAGATAAATAAAT 1077
QY 1570 GTCTGTGAGTGTGTAATTTAAACAAATTTACATTTGATCTGGAATTTTCTGTGAATTT 1629
Db 1078 GTCTGTGAGTGTGTAATTTTACCAAACTTACATTTGATCTGGAATTTTCTGTGAATTT 1137
QY 1630 AGTGAAACAGTCACCATGTGTTCTTCAAGATCTCTGTTACAAACCACTTTCCTGGTGGAT 1689
Db 1138 AGTGAAACAGTCACCATGTGTTCTTCAAGATCTCTGTTACAAACTACTTTTCTCGTGGAT 1197
QY 1690 AACAAAAGGTCTTTTGAACCTCATCTCATCAAGACATGTTGAAGATGACATTCGCTCT 1749
Db 1198 AACAAAAGGTCTTTTGAACCTCATCTCATCAAGACATGTTGAAGATGACATTCGCTCT 1257
QY 1750 TTTGTGAGATTCCTCGAGTGTCTTTCCCAAGTGTCTACCTATATAATTAATCACAGGCTA 1809
Db 1258 TTTGTGAG--TCCTCGGTGCTCTCTCCCAAGTGTGCTGTTATATAATTAATCACAGGCTA 1315

QY 1810 AGGACTGTATCGACTCCTTTGTTTACTCACTGTGTTGCGCCATCTGTAGTCTTATTTCAGA 1869
Db 1316 AGGACTGTATCGACTCCTTTGTTTACTCACTGTGTTGCGCCATCTGTAGTCTTATTTCAGA 1375
QY 1870 TCCATGACATATAACAGGCTCGACAGAGATAGCTTGTGTATCTTCTTGAGAAATTTG 1929
Db 1376 TCCATGACATATAACAGGCTCGACAGAGATAGCTTGTGTATCTTCTTGAGAAATTTG 1435
QY 1930 CCACCTTGCAGGATGAGTTTATGACATTTTATTAATAGCAGAGAGGTTGATGACG 1989
Db 1436 CCACCTTGCAGATG-----AGGACAGAGAGGTCGATGACG 1472
QY 1990 GCTTTCACCATGCTGTTGTAACAGGAAACCCCAAGGCAACATTTGGCCCTGTTAGGTAC 2049
Db 1473 GCTTTCACCATGCTGTTGTAAGCAGGAGCCTCAGCGACAGCATCTGCGCTGTTAGGAAC 1532
QY 2050 CTGGTCTCTTACCATAACTTGCATTTATGATACAGTACTTCTTAAGTGGCTTTGATTT 2109
Db 1533 TTGGTCTCTTACCATAGCCTTGGGATTTATGATCCAGTACTGCTCAGCGGCTTTGATCT 1592
QY 2110 GGAACTCTACAGTATGACGAGTACTATTATATATATATGTTGTTATCTCTCAATTCCTTTA 2169
Db 1593 GGAGCTCTACAGCATGACGAGTACTATTATATCTACTGTTACTCTCTGAGTCTCTGTA 1652
QY 2170 CGCATGTTGATGTTCAACATTTGATGCTGCGATGGCTCTCAATGGCAGAGAAAGAT 2229
Db 1653 CGCATGTTGATGTTCAACCTTGGCCGTGCTGACGGCTCTCAGATGGCAGAGAGAGAT 1712
QY 2230 AATGGAAGACGACAGAAAGCGTGTAGTTAAATAAACAAGAAAGAAAGAAAGTTG 2289
Db 1713 AATGGAAGACGACAGAAAGCGCGCAGCAAGAAACAAGAAAGAAAGAAAGTTG 1772
QY 2290 CCCATTGAGCGAGATCAATGAGCCAGCATATCAGAAATGTGTGCTGGAATGTT 2349
Db 1773 TCCATTGAGCGAGATCAATGAGCCAGCCTATCAGAAATGTGTGCTGGATGTT 1832
QY 2350 TAAAAACATGTTGATTTGACATGAGCGCAAGTACGTAAACCGAAGTTTGAGCTTGA 2409
Db 1833 CAAAAACATGTTGATTTGACATGAGTGGCAAGTGGCAAGCGCCAAATTTGAGCTTGA 1892
QY 2410 TAGTGAACAGTTCCGTGTATGAAACACAGGTTTGTCCATTCACAGTGTGATGACCCGCC 2469
Db 1893 CAGTGAACAGTTCCGTGTATGACAGGTTTGTCCGTTCACAGCGTGTGATGACCCGCC 1952
QY 2470 GCCAGTGCACTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAGCCCTCTCC 2529
Db 1953 ACCAGTGCACTACTTACAGTTCAAGGAAATGTCTGACCTCAGTAATATAGCCCTCTCC 2012
QY 2530 TCAGTCTCTGAACTGTATGTCGACGTAGTAAGCATCTTTCACAGGCAAAATGATTT 2589
Db 2013 TCAGCCTCCAGAGCTGTATGTCGACGTAGTAAGCATCTTTCAGAGGCAAAATGATCTCT 2072
QY 2590 GGAATAATTTCTTAAACCCGACCATGAGTTTATAGAAATTTTAAAGTTGCCAAACCCAA 2649
Db 2073 TGAGAGCATCCCAATGTCAGACCGGAGGTTCAGAGATCTTAAAGTTGCCAAAGCTTAA 2132
QY 2650 CTTTGTGTTATGAAGTTTATGTCAGGAGGACACAAAAGGATCTTAAAGTTCTCTCTGA 2709
Db 2133 CTTTGTGTTATGAAGCTCTTACAGGAGGACACAAAGAGAGTCAAGAGTTCTCTCTGA 2192
QY 2710 ATTTGAATTTCTGCTCATTAATATTTTCTGTTGTAACCTTGTGTAAGAGATCTGG 2769
Db 2193 ATTTGAATTTCTGTTTCAAAAATCTTCTGTTGTAACCTTGTGTAAGAGATCTGA 2252
QY 2770 GAGTGGCCATAAAGGGCAGAGTCTCTTTCAGACCCAACTCTTAGAGGGCAGATCACC 2829
Db 2253 AGATGACCAT-----GTCTACTCTGGGTACACCA 2283
QY 2830 AGGCTCCACATCACGGGAAGTGAGATGATTTCTTGGGTAAACAACATCTATTAAGCA--- 2886
Db 2284 GTGTGCCACACCGGTGACATAAAGTGGCTCTCTTGAATGACAGCTTGTGTAAGAGAGAC 2343
QY 2887 -ATATTTTAGTTTGAACGCTTTATATGACATGAATGAATACTGTGTTTAAAGTGGTT 2945

Db 2344 TTCAGTATCAGTACACACCCCTTACCTGGAGAGAAG-AACTGCTGTTTAAAGTGTT 2402
Qy 2946 TATTATGTTCCATGGAAGAACTGGTCTTATTGAATGCAATGATGAACTTATATGTTT 3005
Db 2403 TGTATATCTCATGGGTGTACGGG----CTGGATGCAATGGTGAAGCTTACATGGTTT 2458
Qy 3006 TATTACAGATTAAATCAAAATCAATTTTATGAATGATGAGTGAAATAGTGTATATA 3065
Db 2459 TATTACAGACTTCATGTTAAACTTTTAAATGAATGATGTTGTG-AAATAGTGTGTA 2517
Qy 3066 AAGGTTAATAATTTCTTGACAAAA 3091
Db 2518 AAGGTTAATAATTTCTTGACAAAA 2543

RESULT 7

AB551309

ID AB551309 standard; cDNA; 719 BP.

AC AB551309;

XX

DT 21-OCT-2002 (first entry)

XX

DE cDNA encoding human secretory protein #7.

XX

Human; secretory polypeptide; SPTM; actinic keratosis; arteriosclerosis; bursitis; cirrhosis; hepatitis; polycythemia vera; anaemia; psoriasis; primary thrombocytopenia; cancer; adenocarcinoma; leukaemia; myeloma; sarcoma; immune system disorder; acquired immunodeficiency syndrome; AIDS; allergy; asthma; Crohn's disease; diabetes mellitus; gout; glomerulonephritis; Goodpasture's syndrome; thyroiditis; pancreatitis; hepatitis; multiple sclerosis; osteoporosis; Reiter's syndrome; rheumatoid arthritis; neurological disorder; epilepsy; stroke; dementia; Alzheimer's disease; Pick's disease; Huntington's disease; mood; anxiety; Parkinson's disease; central nervous system disorder; mental disorder; schizophrenic disorder; amnesia; Tourette's disorder; transgenic animal; gene therapy; gene; ss.

XX

OS Homo sapiens.

XX

WO200257304-A2.

XX

PD 25-JUL-2002.

XX

PF 15-JAN-2002; 2002WO-US001340.

XX

PR 16-JAN-2001; 2001US-0261864P.

XX

PR 16-JAN-2001; 2001US-0261865P.

XX

PR 16-JAN-2001; 2001US-0261979P.

XX

PR 16-JAN-2001; 2001US-0261981P.

XX

PR 17-JAN-2001; 2001US-0262164P.

XX

PR 17-JAN-2001; 2001US-0262208P.

XX

PR 17-JAN-2001; 2001US-0262313P.

XX

PR 19-JAN-2001; 2001US-0262599P.

XX

PR 19-JAN-2001; 2001US-0262760P.

XX

PR 19-JAN-2001; 2001US-0263063P.

XX

PR 19-JAN-2001; 2001US-0263066P.

XX

PR 19-JAN-2001; 2001US-0263069P.

XX

PR 19-JAN-2001; 2001US-0263070P.

XX

PR 19-JAN-2001; 2001US-0263074P.

XX

PR 19-JAN-2001; 2001US-0263076P.

XX

PR 19-JAN-2001; 2001US-0263077P.

XX

PR 19-JAN-2001; 2001US-0263329P.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

XX

PI Panzer SR, Lincoln SE, Altus CM, Dufour GE, Hillman JL, Jones AL;

PI Dam IC, Liu TF, Harris B, Flores V, Daffo A, Marwaha R, Chen AJ;

PI Chang SC, Gerstein EH, Peralta CH, David MH, Lewis SA;

XX

XX

DR WPI; 2002-590716/63.

XX

DR P-PSDB; ABG69816.

XX

XX New purified secretory polypeptides and polynucleotides, useful in the
PT diagnosis, study, prevention or treatment of diseases associated with
PT decreased expression of functional secretory molecules, e.g. AIDS, cancer
PT or allergies.
XX
PS Claim 1; Page 247; 340pp; English.

XX The invention describes an isolated polynucleotide a naturally occurring
XX polynucleotide sequence at least 90 % identical to it, a polynucleotide
XX complementary to it or an RNA equivalent of it, the purified secretory
XX polypeptides (SPTM) and polynucleotides are useful in the diagnosis,
XX study, prevention or treatment of diseases associated with decreased
XX expression of functional SPTM, e.g. actinic keratosis, arteriosclerosis,
XX bursitis, cirrhosis, hepatitis, polycythemia vera, primary
XX thrombocytopenia, anaemia, psoriasis, cancers including adenocarcinoma,
XX leukaemia, myeloma or sarcoma, immune system disorder such as acquired
XX immunodeficiency syndrome (AIDS), allergies, asthma, Crohn's disease,
XX diabetes mellitus, glomerulonephritis, Goodpasture's syndrome, gout,
XX Hashimoto's thyroiditis, hepatitis, multiple sclerosis, osteoporosis,
XX pancreatitis, Reiter's syndrome, autoimmune thyroiditis or rheumatoid
XX arthritis, neurological disorders such as epilepsy, stroke, Alzheimer's
XX disease, Pick's disease, Huntington's disease, dementia, Parkinson's
XX disease, other developmental disorder of the central nervous system,
XX mental disorder including mood, anxiety or schizophrenic disorder,
XX amnesia or Tourette's disorder. The polynucleotides may be used in
XX hybridisation and amplification technologies, e.g. in assessing gene
XX expression patterns, to develop a transcript image for a particular cell
XX or tissue, or to create transgenic animals to model human disease. This
XX sequence encodes a human secretory protein isolated in the invention

XX Sequence 719 BP; 223 A; 113 C; 188 G; 195 T; 0 U; 0 Other;

Query Match 23.2%; Score 717.4; DB 6; Length 719;
Best Local Similarity 99.9%; Pred. No. 5.8e-171;
Matches 718; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 474 GCGCGCGCCGAGCGCGTCTGTTATTTCCGTGGTCCGACAGTCGCGCGCGCGGT 533

Db 1 GCGCGCGCCGAGCGCGTCTGTTATTTCCGTGGTCCGACAGTCGCGCGCGCGGT 60

Qy 534 GACCACGGGAGAGTAGGCATATGTTTATGAAAGCTTCTGTAGATGATGAGATTACAG 593

Db 61 GACCACGGGAGAGTAGGCATATGTTTATGAAAGCTTCTGTAGATGATGAGATTACAG 120

Qy 594 ATGGGAGCTCAGTATGCCAGAAAAATGAGAAAAAGCAATACAACTGGTGGACATTAC 653

Db 121 ATGGGAGCTCAGTATGCCAGAAAAATGAGAAAAAGCAATACAACTGGTGGACATTAC 180

Qy 654 CCAAGATTTTCAAGAAGCTTCTCGAGAAATTAAGTTGGGAGAACTACTTTCATGTAAGCT 713

Db 181 CCAAGATTTTCAAGAAGCTTCTCGAGAAATTAAGTTGGGAGAACTACTTTCATGTAAGCT 240

Qy 714 ATTTGGTCTTTTGAAGCCATGCTGCTGTTATGAAATGATGATGCCAAGATGATGCTGG 773

Db 241 ATTTGGTCTTTTGAAGCCATGCTGCTGTTATGAAATGATGATGCCAAGATGATGCTGG 300

Qy 774 CATGATTGGAACCAAGTTAATCGAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGG 833

Db 301 CATGATTGGAACCAAGTTAATCGAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGG 360

Qy 834 CACTATTAAATTAAGATCTCACCTTCCTGAACTGATAGGATTTATGGATACATGTTT 893

Db 361 CACTATTAAATTAAGATCTCACCTTCCTGAACTGATAGGATTTATGGATACATGTTT 420

Qy 894 TTGCTGTTTGTATACGCTGGTTAGAGGCCATTCTACTGGCACAGACAGTATTTACGCTCT 953

Db 421 TTGCTGTTTGTATACGCTGGTTAGAGGCCATTCTACTGGCACAGACAGTATTTACGCTCT 480

Qy 954 TTACATTTCAATCCAGACTTTATAGAGATCTCTGCTATGAGGCTTTTGTCTCTGGGAT 1013

Db 481 TTACATTTCAATCCAGACTTTATAGAGATCTCTGCTATGAGGCTTTTGTCTCTGGGAT 540

QY 1014 CTTGAAATCTGTGACATTCAGGGGAAAGTAATAAAGCTGCTGTTTTTGAAGAGGA 1073
 Db |||||||
 541 CTTGAAATCTGTGACATTCAGGGGAAAGTAATAAAGCTGCTGTTTTTGAAGAGGA 600
 QY 1074 AGATTTTCAGTCAATGACATTCATGATTTAAATATGGCTTAAACAGTGTGACAGATCTTCGAGT 1133
 Db |||||||
 601 AGATTTTCAGTCAATGACATTCATGATTTAAATATGGCTTAAACAGTGTGACAGATCTTCGAGT 660
 QY 1134 TACAGGATGCTAAAGATGTGAGGATGACATGCAAGAAAGAGTAAAGAGTACTCGAA 1192
 Db |||||||
 661 TACAGGATGCTAAAGATGTGAGGATGACATGCAAGAAAGAGTAAAGAGTACTCGAA 719

RESULT 8

ACH33702
 ID ACH33702 standard; cDNA; 482 BP.

XX ACH33702;

XX 13-OCT-2003 (first entry)

DE Human endothelial cell cDNA #1835.

XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
 KW genome mapping; biodiversity; genetic disorder.

XX Homo sapiens.

XX US2003073623-A1.

PN 17-APR-2003.

XX 30-JUL-2001; 2001US-00918995.

XX 30-JUL-2001; 2001US-00918995.

XX (DRMA/) DRMANAC R T.

PA (LABA/) LABAT I.

PA (STAC/) STACHE-CRAIN B.

PA (DICK/) DICKSON M C.

PA (JONE/) JONES L W.

XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;

XX WPI; 2003-615964/S9.

XX New polynucleotide sequences obtained from various cDNA libraries, useful
 PT as hybridization probes, as oligomers for PCR, for chromosome and gene
 PT mapping, in the recombinant production of protein, or in generating
 PT antisense DNA or RNA.

XX Claim 1; SEQ ID NO 20914; 44pp; English.

XX The invention relates to an isolated polynucleotide comprising any one of
 CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
 CC determined by the technique of SBH (sequencing by hybridisation). Also
 CC included is a purified polypeptide comprising a sequence corresponding to
 CC a reading frame of the novel polynucleotide. The nucleic acid sequences
 CC are useful in diagnostics as expressed sequence tags (EST) for
 CC identifying expressed genes or for physical mapping of the human genome,
 CC in forensics, in assessing biodiversity, or in identifying mutations
 CC responsible for genetic disorders and other traits. The nucleotide
 CC sequences are also useful as hybridisation probes, as oligomers for PCR,
 CC for chromosome and gene mapping, in the recombinant production of
 CC protein, or in generating antisense DNA or RNA. The purified polypeptide
 CC is useful for generating antibodies specific for it. The present sequence
 CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?DocID=20030073623

XX Sequence 482 BP; 145 A; 82 C; 124 G; 130 T; 0 U; 1 Other;

Query Match 14.8%; Score 451.2; DB 8; Length 482;
 Best Local Similarity 99.3%; Pred. No. 1.1e-103;
 Matches 453; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 491 CGTCGTTATTCGTCGCGACACTGCGTGGCGGGGGTGACACCGGAGAGTAG 550
 Db |||||||
 26 CGAGCTTATTCGTCGTCGCGACAGTGGCTGGCGGGGGTGACACCGGAGAGTAG 85
 QY 551 GCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTGAGGATGGAGCTCAGTATGC 610
 Db |||||||
 86 GCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTGAGGATGGAGCTCAGTATGC 145
 QY 611 CAGAAAAATGGAGAAAGCAATACAAACTGGGTGGACATTCACCAAGATTTTGAAGAG 670
 Db |||||||
 146 CAGAAAAATGGAGAAAGCAATACAAACTGGGTGGACATTCACCAAGATTTTGAAGAG 205
 QY 671 CTTGTGAGAAATTTAAAGTTTGGGAGAACTACTTTCATGATAAGCTATTGTCCTTTTGAAG 730
 Db |||||||
 206 CTTGTGAGAAATTTAAAGTTTGGGAGAACTACTTTCATGATAAGCTATTGTCCTTTTGAAG 265
 QY 731 CCATGTCCTATTTGAAATGATGGATCCCAAGATGGATGCGCATGATGGAAACCAAG 790
 Db |||||||
 266 CCATGTCCTATTTGAAATGATGGATCCCAAGATGGATGCGCATGATGGAAACCAAG 325
 QY 791 TTAATCGAAAAGTCTCAATTTTGAACAAGCTATCAAGGATGGCATATTAATAATTAAAG 850
 Db |||||||
 326 TTAATCGAAAAGTCTCAATTTTGAACAAGCTATCAAGGATGGCATATTAATAATTAAAG 385
 QY 851 ATCTCACCCTTGCTGAACCTGATAGGATATGGATACATGTTTTCCTGCTGTTGATAACGT 910
 Db |||||||
 386 ATCTCACCCTTGCTGAACCTGATAGGATATGGATACATGTTTTCCTGCTGTTGATAACGT 445
 QY 911 GGTAGAAGGCCATTCACCTGGCACACAGATATTTA 946
 Db |||||||
 446 GGTAGAAGGCCATTCACCTGGCACACAGATATTTA 481

RESULT 9

ADC31969

ID ADC31969 standard; cDNA; 482 BP.

XX ADC31969;

XX 18-DEC-2003 (first entry)

XX Human novel cDNA contig sequence, SEQ ID NO:2051.

XX Human; diagnostic; drug screening; forensics; gene mapping;
 KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;
 KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
 KW ulcers; osteoporosis; autoimmune disease; cancer;
 KW molecular weight marker; food supplement; antiparkinsonian; nootropic;
 KW neuroprotective; antianemic; anticoagulant; thrombolytic; vulnerary;
 KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;
 KW gene therapy; chromosome 9; ss.

XX Homo sapiens.

XX WO2003029271-A2.

XX 10-APR-2003.

XX 24-SEP-2002; 2002WO-US030474.

XX 24-SEP-2001; 2001US-0324631P.

XX (HYSE-) HYSEQ INC.

XX Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
 PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
 PI Haley-Vicente D, Drmanac RT;

XX WPI; 2003-371981/35.

DR P-PSDB; ADC32736.
XX New polynucleotide and polypeptide useful for diagnosing, preventing or
PT treating conditions such as neurodegenerative diseases, anemias, platelet
PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
PT cancer.
XX
XX Example 2; SEQ ID NO 2051; 1185pp; English.
XX
XX The invention relates to 971 novel human cDNA sequences (ADC29919-
CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The
CC invention also relates to nucleic acid sequences over 99% identical with
CC the novel human cDNAs. The invention additionally encompasses expression
CC vectors and host cells comprising a nucleic acid of the invention; the
CC recombinant production of a polypeptide of the invention; an antibody
CC against a polypeptide of the invention; a method of detecting
CC polynucleotides or polypeptides of the invention; and methods of
CC identifying a compound which binds to a polypeptide of the invention. The
CC invention further discloses methods of preventing, treating or
CC ameliorating a medical condition; kits comprising polynucleotide probes
CC and/or monoclonal antibodies for carrying out the methods of the
CC invention; methods for the identification of compounds that modulate the
CC expression or activity of the polynucleotide and/or polypeptide; and 767
CC contig sequences corresponding to the cDNA sequences of the invention
CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
CC -ADC33394). The nucleic acids and polypeptides of the invention are
CC useful in diagnostics, drug screening, forensics, gene mapping, in the
CC identification of mutations responsible for genetic disorders or other
CC traits, for assessing biodiversity, and in producing many other types of
CC data and products dependent on DNA and amino acid sequences. They are
CC also used for treating diseases such as Parkinson's disease, Alzheimer's
CC disease and other neurodegenerative diseases, anaemia, platelet
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
CC cancer. The nucleic acids may also be used as hybridisation probes or
CC primers, and in the recombinant production of a protein. The polypeptides
CC are also useful in generating antibodies, as molecular weight markers,
CC and as food supplements. The present sequence represents a human contig
CC sequence used in an example of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX
XX Sequence 482 BP; 145 A; 82 C; 124 G; 130 T; 0 U; 1 Other;
Query Match 14.6%; Score 451.2; DB 9; Length 482;
Best Local Similarity 99.3%; Pred. No. 1.1e-103;
Matches 453; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 491 CGTCGTTATTTCCTGCTCCGACAGTCGCGCGCGCGGTGACACCGGAGAGTAG 550
DB 26 CGAGCTTATTTCCTGCTCCGACAGTCGCGCGCGCGGTGACACCGGAGAGTAG 85
QY 551 GCATAATGGTTATGAAGCTTCTTAGATGATGACGATTCAGGATGGAGTCTAGTATGC 610
DB 86 GCATAATGGTTATGAAGCTTCTTAGATGATGACGATTCAGGATGGAGTCTAGTATGC 145
QY 611 CAGAAAAATGAGAAAAAGCAATACAACTGGGTGGACATTCACCAAGATTTTCAAGAAG 670
DB 146 CAGAAAAATGAGAAAAAGCAATACAACTGGGTGGACATTCACCAAGATTTTCAAGAAG 205
QY 671 CTTGTCGAGAAATTAAGTTGGGAGAACTACTTCATGATTAAGCTATTGGTCTTTTGAAG 730
DB 206 CTTGTCGAGAAATTAAGTTGGGAGAACTACTTCATGATTAAGCTATTGGTCTTTTGAAG 265
QY 731 CCATGTCGCTATTGAATGATGATCCAGATGGATGCTGGCATGATTTGGAACCAAG 790
DB 266 CCATGTCGCTATTGAATGATGATCCAGATGGATGCTGGCATGATTTGGAACCAAG 325
QY 791 TTAATCGAAAAAGTTCTCAATTTTGAACAAAGCTATCAAGATGGCAGCTATTAAAAATTAAG 850
DB 326 TTAATCGAAAAAGTTCTCAATTTTGAACAAAGCTATCAAGATGGCAGCTATTAAAAATTAAG 385
QY 851 ATCTCACCTTGCCTGAACTGATAGGATATTGGATACATGTTTTTGTGTTGTAACGT 910

DB 386 ATCTCACCTTGCCTGAACTGATAGGATATTGGATACATGTTTTTGTGTTGTAACGT 445
QY 911 GGTTAGAAGCCCAATTCATCTGCGACAGACAGATATTTA 946
DB 446 GGTTAGAAGCCCAATTCATCTGCGACAGACAGATATTTA 481

RESULT 10
ABX50281
ID ABX50281 standard; cDNA; 398 BP.
XX
XX AC ABX50281;
XX
XX DT 25-FEB-2003 (first entry)
XX
XX Bovine EST associated with lactation/muscle/fat deposition #210.
XX
XX Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
XX muscle deposition; fat deposition; genome mapping; gene identification;
XX gene analysis; cattle breeding.
XX
XX Bos Taurus.
XX
XX US2002137160-A1.
XX
XX 26-SEP-2002.
XX
XX 26-OCT-2001; 2001US-00983965.
XX
XX 17-DEC-1998; 98US-0113678P.
XX
XX 15-DEC-1999; 99US-00465231.
XX
XX (BYAT// BYATT J C.
XX (MATH// MATHIALAGAN N.
XX (TAON// TAO N.
XX (WARR// WARREN W C.
XX
XX Byatt JC, Mathialagan N, Tao N, Warren WC;
XX
XX WPI; 2003-102386/09.
XX
XX Purified nucleic acid molecules, useful for genome mapping, gene
XX identification and analysis, cattle breeding or preparation of constructs
XX for cattle gene expression and genetically improved cattle.
XX
XX Claim 2; SEQ ID NO 210; 38pp; English.

XX The invention relates to a purified nucleic acid molecule associated with
XX lactation or muscle and fat deposition (designated LMFD), derived from
XX cattle, and the LMFD nucleic acid can specifically hybridise to a second
XX nucleic acid molecule comprising any of 5912 nucleotide sequences,
XX appearing as ABX50072-ABX55983, or complements of them. Also included are
XX ; (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
XX acid linked to a promoter and a 3' non-translated sequence that
XX functions in the cell to cause termination of transcription and addition
XX of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
XX (2) determining a level or pattern of a molecule in a bovine cell or
XX tissue comprising: (a) incubating a marker nucleic acid (comprising any
XX of the 5912 nucleic acid sequences or its complement or fragment) with a
XX complementary nucleic acid molecule obtained from the bovine cell or
XX tissue, where hybridisation between the marker nucleic acid and the
XX complementary nucleic acid permits the detection of the molecule; and (b)
XX detecting the level or pattern of the complementary nucleic acid, where
XX the detection of the complementary nucleic acid is predictive of the
XX level or pattern of the molecule. The LMFD nucleic acid is used for
XX determining a level or pattern of a molecule in a bovine cell or tissue.
XX It is useful for genome mapping, gene identification and analysis, cattle
XX breeding, preparation of constructs for use in cattle gene expression, or
XX for genetically improving cattle. The present sequence is one of the 5912
XX bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The present
XX sequence was not shown in the specification but was obtained in
XX electronic format from the USPTO web site.

KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 XX SNP; cell differentiation; ds.
 OS Homo sapiens.
 XX WO200218632-A2.
 XX 07-MAR-2002.
 XX 01-SEP-2001; 2001WO-EP010074.
 XX 01-SEP-2000; 2000DE-01043826.
 XX 05-SEP-2000; 2000DE-01044543.
 XX (EPIG-) EPIGENOMICS AG.
 XX Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX WPI; 2002-371829/40.
 XX Determining the degree of cytosine methylation in genomic DNA, useful for
 PT diagnosis and prognosis, comprises selective hybridization of amplicons
 PT from chemically treated DNA.
 XX Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention
 XX
 XX Sequence 799 BP; 118 A; 95 C; 297 G; 289 T; 0 U; 0 Other;
 Query Match 11.1%; Score 343.8; DB 6; Length 799;
 Best Local Similarity 83.5%; Pred. No. 2.3e-76;
 Matches 390; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
 QY 87 TCGTCGTTCTGTCGCAACCACTAAGGTCTACGCAACCTCCACGGTTCTCTCCGGCT 146
 DB 2 TGTTCGTTCTGTCGTAATTAATTAAGGTTCAGTAATTTTACGGTTTTTTTCGTTT 61
 QY 147 TCGGTCACCTTCTAGAAATTCACAGGGCGGCGAGAGCGGGCGGCTCTGAGACT 206
 DB 62 TCGGGTTATTTTTTAAAGAAATTTTACAGGGTAGCGTAGACGGGGCGGTTTGAGATT 121
 QY 207 CCGGGCTCCGGCTCTTTCCGGGAACCGCCACTCCAGGACTCCGACAGAGGGTGAAGA 266
 DB 122 TCGGGTTCGTTTTTTTCGGGAATCGTTTATTTAGGATTTTCGATAGAGGGTGAAGA 181
 QY 267 AAGATAACTTCGGTCTCGCGATCGTCTTAATTCGCGAGAGAGAGCGCGCGCCAT 326
 DB 182 AAGATAATTTTCGGTTTCGCGATCGTTTTTAATTTTCGCGAGAGAGAGCGCGTCTTAT 241
 QY 327 CGGCCGAACGAGCGGTGCGGAGGAGGGGTGTGGCCGGGAGCGCGAATGCCCGGG 386
 DB 242 CGGTCGAACGAGGCGGTGGGAGGAGGGGTGTGGTCGGGGAGCGCGAAGTTTTCGG 301

QY 387 AGTAAGGGAGAGGGGGGGTCCGGCTCCCGGCATACGCATCGCTGACGCTCCGG 446
 DB 302 AGTAAGGGAGAGGGGGGGTCCGGCTCCCGGCATACGCATCGCTGACGCTCCGG 361
 QY 447 TCGGGCTGGGCTGAGAGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 506
 DB 362 TCGGGTTGGGTTGAGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 421
 QY 507 GTCGGACAGTGCCTGGCGCGGGTGACCAACCGGAGAAAGTAGCA 553
 DB 422 GTTCGATAGTGCCTGGCGCGGGGTGATTACGGAGAGTAGGTA 468

RESULT 13
 ABV15331/c
 ID ABV15331 standard; cDNA; 356 BP.
 XX AC ABV15331;
 XX 13-SEP-2002 (first entry)
 XX Human prostate expression marker cDNA 15322.
 XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX Homo sapiens.
 XX WO200160860-A2.
 XX 23-AUG-2001.
 XX 20-FEB-2001; 2001WO-US005171.
 XX 17-FEB-2000; 2000US-0183319P.
 XX 16-MAR-2000; 2000US-0189862P.
 XX 25-MAY-2000; 2000US-0207454P.
 XX 09-JUN-2000; 2000US-0211314P.
 XX 18-JUL-2000; 2000US-0219007P.
 XX 13-DEC-2000; 2000US-0255281P.
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX Schlegel R, Endege WO, Monahan JE;
 XX WPI; 2001-662795/76.

Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.
 Claim 1; Page 2572; 11750pp; English.
 The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the efficacy progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
 Sequence 356 BP; 97 A; 80 C; 85 G; 94 T; 0 U; 0 Other;

Query Match 9.4%; Score 292.4; DB 5; Length 356;
 Best Local Similarity 90.5%; Pred. No. 1.5e-63;
 Matches 344; Conservative 0; Mismatches 11; Indels 25; Gaps 2;


```
XX Schlegel R, Endege WO, Monahan JE;
PI WPI; 2001-662795/76.
DR Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX Claim 1; Page 7483; 11750pp; English.
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 439 BP; 113 A; 104 C; 110 G; 112 T; 0 U; 0 Other;

Query Match          9.3%; Score 288.8; DB 5; Length 439;
Best Local Similarity 91.4%; Pred. No. 1.4e-62;
Matches 338; Conservative 0; Mismatches 7; Indels 25; Gaps 2;

QY 1687 GATAACAAAAGGCTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACCTCGG 1746
Db      |||||||
QY 439  GATAACAAAAGGCTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACCTCGG 380
Db      |||||||
QY 1747 TCTTTTGTGCATCCTCCGAGTGTTCCTCCCAAGTGTACCTATATAATAATCACCAGG 1806
Db      |||||||
QY 379  TCTTTTGTGCAG--TCCTCCGGTGTCTTCCCAAGTGTACCTATATAATAATCACCAGG 322
Db      |||||||
QY 1807 CTAAGGACTGTATCGACTCCTTTGTTACTACTGTGTTCGGCCATCTGTAGTCTTTATTC 1866
Db      |||||||
QY 321  CTAAGGACTGTATCGACTCCTTTGTTACTACTGTGTTCGGCCATCTGTAGTCTTTATTC 262
Db      |||||||
QY 1867 AGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTTCTTGAGGAAT 1926
Db      |||||||
QY 261  AGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTTCTTGAGGAAT 202
Db      |||||||
QY 1927 TTGCCACCTTGCAGGATGAGTTTATGACATTTTATTTTATAGGACAGAGAGGTTGATGC 1986
Db      |||||||
QY 201  TTGCCACCTTGCAGGATG-----AGGCAGAGAGAGGTTGATGC 165
Db      |||||||
QY 1987 AGCGCTTCACACCATGCTGTTGAAACAGAGAACCCCAAGGCAACATTTGGCCTGTTTAGG 2046
Db      |||||||
QY 164  AGCGCTTCACACCATGCTGTTGAAACAGAGAACCCCAAGGCAACATTTGGCCTGTTTAGG 105
Db      |||||||
QY 2047 TACCTGGGTC 2056
Db      |||||||
QY 104  TACCTGGGCC 95
```

Search completed: August 10, 2004, 21:54:14
Job time : 1247.45 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 10, 2004, 21:58:32 ; Search time 1436.92 Seconds
(without alignments)
10571.835 Million cell updates/sec

Title: US-10-001-857-42

Perfect score: 3096

Sequence: 1 ttctccacgaactccagg.....attcttgacaaaaaa 3096

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3225727 seqs, 2453303834 residues

Total number of hits satisfying chosen parameters: 6451454

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PTCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PTCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3096	100.0	3096	14	US-10-001-857-42
2	2417.2	78.1	2488	17	US-10-467-433-23
3	2228.4	72.0	2610	10	US-09-814-353-20725
c	4	560.8	18.1	745	10 US-09-814-353-15927
5	509.4	16.5	610	10	US-09-814-353-16155
c	6	496.2	16.0	602	10 US-09-814-353-3218
7	496.2	16.0	602	10	US-09-814-353-9543
8	431.2	14.6	482	10	US-09-918-995-20914
9	434	14.0	434	10	US-09-814-353-3455
10	434	14.0	434	10	US-09-814-353-9771
11	350.8	11.3	398	9	US-09-983-965-210
12	302.6	9.8	345	13	US-10-085-783A-43407
13	302.6	9.8	345	16	US-10-242-535A-43407
14	268.6	8.7	501	9	US-09-783-590-5693

C 15	182	5.9	489	13	US-10-027-632-323193	Sequence 323193,
C 16	182	5.9	489	16	US-10-027-632-323193	Sequence 323193,
C 17	140.4	4.5	151	14	US-10-001-857-41	Sequence 41, Appl
C 18	87.8	2.8	815	14	US-10-001-857-111	Sequence 111, App
20	81.4	2.6	430	13	US-10-085-783A-54186	Sequence 54186, A
20	81.4	2.6	430	16	US-10-242-535A-54186	Sequence 54186, A
21	79.2	2.6	962	13	US-10-424-599-47623	Sequence 47623, A
22	64.2	2.1	2121	17	US-10-437-963-44335	Sequence 44335, A
23	60.8	2.0	456	17	US-10-437-963-51928	Sequence 51928, A
24	60.2	1.9	629	17	US-10-021-323-9375	Sequence 9375, Ap
25	60	1.9	60	10	US-09-908-975-16495	Sequence 16495, A
C 26	59.8	1.9	492	13	US-10-424-599-96349	Sequence 96349, A
C 27	59.6	1.9	852	17	US-10-437-963-68188	Sequence 68188, A
C 28	58.6	1.9	420	17	US-10-021-323-5998	Sequence 5998, Ap
29	58.2	1.9	3163	15	US-10-017-161-1857	Sequence 1857, Ap
30	58.2	1.9	3163	16	US-10-292-798-1513	Sequence 1513, Ap
C 31	58	1.9	925	17	US-10-437-963-44536	Sequence 44536, A
C 32	58	1.9	956	13	US-10-027-632-31508	Sequence 31508, A
C 33	58	1.9	956	16	US-10-027-632-31508	Sequence 31508, A
C 34	57.8	1.9	809	13	US-10-424-599-26988	Sequence 26988, A
C 35	57.4	1.9	497	13	US-10-424-599-133788	Sequence 133788, A
C 36	56.6	1.8	2432	17	US-10-437-963-64501	Sequence 64501, A
C 37	56.4	1.8	580	13	US-10-424-599-2320	Sequence 2320, Ap
C 38	56.2	1.8	1211	17	US-10-437-963-27585	Sequence 27585, A
39	56.2	1.8	5452	15	US-10-017-161-1481	Sequence 1481, Ap
40	56.2	1.8	5452	16	US-10-292-798-1189	Sequence 1189, Ap
41	56	1.8	451	13	US-10-240-425-721	Sequence 721, App
C 42	56	1.8	630	17	US-10-437-963-18774	Sequence 18774, A
43	56	1.8	1117	15	US-10-017-161-1403	Sequence 1403, Ap
44	56	1.8	1117	16	US-10-292-798-1141	Sequence 1141, Ap
C 45	55.8	1.8	424	17	US-10-437-963-49515	Sequence 49515, A

ALIGNMENTS

RESULT 1

US-10-001-857-42

; Sequence 42, Application US/10001857

; Publication No. US20020183500A1

; GENERAL INFORMATION:

; APPLICANT: Macina, Roberto

; APPLICANT: Recipon, Herve

; APPLICANT: Chen, Sei-yu

; APPLICANT: Sun, Yongming

; APPLICANT: Liu, Chenghua

; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Pro

; FILE REFERENCE: DEX-0273

; CURRENT APPLICATION NUMBER: US/10/001.857

; CURRENT FILING DATE: 2001-11-20

; PRIOR APPLICATION NUMBER: 60/252,054

; PRIOR FILING DATE: 2000-11-20

; NUMBER OF SEQ ID NOS: 208

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 42

; LENGTH: 3096

; TYPE: DNA

; ORGANISM: Homo sapien

; US-10-001-857-42

Query Match 100.0%; Score 3096; DB 14; Length 3096;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 3096; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTCTCTACGAAATCCCGAGCGCTGTATAGAAACATAAAATCGTTGTCAGGACGAGTA 60

Db 1 TTCTCTACGAAATCCCGAGCGCTGTATAGAAACATAAAATCGTTGTCAGGACGAGTA 60

Qy 61 GCACGCTGTTGCTCTCGGAGCTTGGCTGCTCGTTCGCAACCACTAAGGTCTACG 120

Db 61 GCACGCTGTTGCTCTCGGAGCTTGGCTGCTCGTTCGCAACCACTAAGGTCTACG 120

Qy 121 CAAACCTCCAGGGTTTCCTCCGCTTCGGCTGCACCTTCTAAGAAATCCGAGGGCA 180

Db 121 CAAACCTCCAGGTTCTTCCGCTTCCGCTCACTTTCTAAGAAATTCACAGAGGCA 180
QY 181 GCGCAGACGGGCGGCTCTGAGCTCCGGCTCCGCTCTTTCCGGGAACCGCCACTA 240
Db 181 GCGCAGACGGGCGGCTCTGAGCTCCGGCTCCGCTCTTTCCGGGAACCGCCACTA 240
QY 241 CCCAGGATCCGACAGAGGCTGAAAGAGATAAATTCGCGTCTCGCGATCGTCTTAATC 300
Db 241 CCCAGGATCCGACAGAGGCTGAAAGAGATAAATTCGCGTCTCGCGATCGTCTTAATC 300
QY 301 TCGCAGAGAGAAAGCGCGCCCATCGCCGGAACGAGCGGCTCGCGAGGAGGGGGTG 360
Db 301 TCGCAGAGAGAAAGCGCGCCCATCGCCGGAACGAGCGGCTCGCGAGGAGGGGGTG 360
QY 361 TGGCGGGGAGCGCAAGTCCCGGGAGTAAGGAGAGGGGGCGGGTCCGCGTCCCGG 420
Db 361 TGGCGGGGAGCGCAAGTCCCGGGAGTAAGGAGAGGGGGCGGGTCCGCGTCCCGG 420
QY 421 GCATACGATCGTGCACGCTGCGCGTCCGCTGAGAGGGGAGGGGGCGCGGCG 480
Db 421 GCATACGATCGTGCACGCTGCGCGTCCGCTGAGAGGGGAGGGGGCGCGGCG 480
QY 481 GCGGAGGCGGCTGTTATTTCCGTGTCGGACAGTGCCTGCGCGCGCGGTGACCAAG 540
Db 481 GCGGAGGCGGCTGTTATTTCCGTGTCGGACAGTGCCTGCGCGCGCGGTGACCAAG 540
QY 541 GGAGAGTAGGCATTAATGGTATGAAGCTCTGTAGATGATGACGATTCAGGATGGAG 600
Db 541 GGAGAGTAGGCATTAATGGTATGAAGCTCTGTAGATGATGACGATTCAGGATGGAG 600
QY 601 CTCAGTATGCCAGAAAAATGGAGAAAAACAATAAACAATGGGTGGCAATACCCAAGAT 660
Db 601 CTCAGTATGCCAGAAAAATGGAGAAAAACAATAAACAATGGGTGGCAATACCCAAGAT 660
QY 661 TTTGAAGAGCTTGTGAGAGATTAAGTTGGAGAACTACTTCAATGAAGCTATTTGGT 720
Db 661 TTTGAAGAGCTTGTGAGAGATTAAGTTGGAGAACTACTTCAATGAAGCTATTTGGT 720
QY 721 CTTTGTGAAGCATGTCTGCTATTGAATGATGGATCCCAAGATGGATGGATGAT 780
Db 721 CTTTGTGAAGCATGTCTGCTATTGAATGATGGATCCCAAGATGGATGGATGAT 780
QY 781 GGAAACCAAGTTAATCGAAAAAGTTCTCAATTTGAAACAAGCTATCAAGGATGGCACTAT 840
Db 781 GGAAACCAAGTTAATCGAAAAAGTTCTCAATTTGAAACAAGCTATCAAGGATGGCACTAT 840
QY 841 AAAATTAAGATCTCACCTTGCCTGAACTGATAGGATATGGATACATGTTTTCGT 900
Db 841 AAAATTAAGATCTCACCTTGCCTGAACTGATAGGATATGGATACATGTTTTCGT 900
QY 901 TTGATAACGTTGTTAGAGGCAATTCAGTGCACAGACAGTATTTACGTCCTTTACAT 960
Db 901 TTGATAACGTTGTTAGAGGCAATTCAGTGCACAGACAGTATTTACGTCCTTTACAT 960
QY 961 CATATCCAGATTTATAGAGATCTGCTATAGAGGCTTTGCTTGGGAATCTTGAAA 1020
Db 961 CATATCCAGATTTATAGAGATCTGCTATAGAGGCTTTGCTTGGGAATCTTGAAA 1020
QY 1021 ATCTGTGACATTCAGGAAAAAGTAAATAAAGCTGCTGTTTGAAGAGGAAGTTTT 1080
Db 1021 ATCTGTGACATTCAGGAAAAAGTAAATAAAGCTGCTGTTTGAAGAGGAAGTTTT 1080
QY 1081 CAGTCAATGACTTATGAGATTTAAATGGCTAACAGTGTGACAGATCTTCAGGTTACAGGC 1140
Db 1081 CAGTCAATGACTTATGAGATTTAAATGGCTAACAGTGTGACAGATCTTCAGGTTACAGGC 1140
QY 1141 ATGCTAAAAGATGTGGAGATGACATGCAAGAGAGATAAGAGTACTCGAAGTCGACAA 1200
Db 1141 ATGCTAAAAGATGTGGAGATGACATGCAAGAGAGATAAGAGTACTCGAAGTCGACAA 1200
QY 1201 GGAGAGAAAGAGATCCAGAGTTGAATAGAACCAACCAATGTTAGCAATTTTACG 1260
Db 1201 GGAGAGAAAGAGATCCAGAGTTGAATAGAACCAACCAATGTTAGCAATTTTACG 1260

Db 1201 GGAGAGAAAGAGATCCAGAGTTGAATAGAACCAACCAATGTTAGCAATTTTACG 1260
QY 1261 AGAGTGAAATTTTACTCGTGTGTTACTGACAGTCTTATAGCCTTTTCTAAGAAAAGAGACC 1320
Db 1261 AGAGTGAAATTTTACTCGTGTGTTACTGACAGTCTTATAGCCTTTTCTAAGAAAAGAGACC 1320
QY 1321 AGTGTGTTGACAGAGCTCAAAAATTTGATGTTCAAGCAGCAGATCTTCTTCTGCCAT 1380
Db 1321 AGTGTGTTGACAGAGCTCAAAAATTTGATGTTCAAGCAGCAGATCTTCTTCTGCCAT 1380
QY 1381 CATATTCATTCATCATGCGCATCCAGGCCAGAGATGATCTACAAAAGGAGATCATCA 1440
Db 1381 CATATTCATTCATCATGCGCATCCAGGCCAGAGATGATCTACAAAAGGAGATCATCA 1440
QY 1441 ATTATGATGGTGTGAAACCCCTTGTGAAACAGAGGCTACTTCCACCTACCTTCCCTCGA 1500
Db 1441 ATTATGATGGTGTGAAACCCCTTGTGAAACAGAGGCTACTTCCACCTACCTTCCCTCGA 1500
QY 1501 TATGCAAAAAATAATTAAGAGGAGAAATGTTGAACTATTTTGCAGATTAATAGATAGA 1560
Db 1501 TATGCAAAAAATAATTAAGAGGAGAAATGTTGAACTATTTTGCAGATTAATAGATAGA 1560
QY 1561 ATAAAACTGTCTGTGAGGTTGTGAAATTTTAACTATTTTACATTTGATCTCGATTTTTC 1620
Db 1561 ATAAAACTGTCTGTGAGGTTGTGAAATTTTAACTATTTTACATTTGATCTCGATTTTTC 1620
QY 1621 TGTGAATTTAGTGAACAGTACCAGTGTCTTTCAGATCTTGTTCACAAACCACTTTC 1680
Db 1621 TGTGAATTTAGTGAACAGTACCAGTGTCTTTCAGATCTTGTTCACAAACCACTTTC 1680
QY 1681 CTGTGGATAACAAAAAGGCTTTGGAACTCATCTCATGCAAGACATGCTGCAAGATGCA 1740
Db 1681 CTGTGGATAACAAAAAGGCTTTGGAACTCATCTCATGCAAGACATGCTGCAAGATGCA 1740
QY 1741 CTTCCGCTTTTGTGAGATCTCCGAGTCTTTCCCAAGTGTCTACCTATATAAATC 1800
Db 1741 CTTCCGCTTTTGTGAGATCTCCGAGTCTTTCCCAAGTGTCTACCTATATAAATC 1800
QY 1801 ACCAGGCTAAGGATGATCGACTCTTTGTTTACTCATCTGTTTGGGCACTTGTAGTC 1860
Db 1801 ACCAGGCTAAGGATGATCGACTCTTTGTTTACTCATCTGTTTGGGCACTTGTAGTC 1860
QY 1861 TTAATTCAGATCCATGGACATAAAGGCTCCAGAGAGATAAGCTTGGTCTATATTTTG 1920
Db 1861 TTAATTCAGATCCATGGACATAAAGGCTCCAGAGAGATAAGCTTGGTCTATATTTTG 1920
QY 1921 AGGAATTTGCCACTTTCAGGATGAGTTTATGACATTTATTTTAAATAGGAGAGAGGT 1980
Db 1921 AGGAATTTGCCACTTTCAGGATGAGTTTATGACATTTATTTTAAATAGGAGAGAGGT 1980
QY 1981 TGATGACGCTTCCACACCTTTCACCACTGCTTTGAAACAGAGAACCCCAAGGCAACATTTGGCCTG 2040
Db 1981 TGATGACGCTTCCACACCTTTCACCACTGCTTTGAAACAGAGAACCCCAAGGCAACATTTGGCCTG 2040
QY 2041 TTTAGGTACCTTGGTCTTTTACCATAACCTTTCGATTTATGATACAGTACCTTCTAAGTG 2100
Db 2041 TTTAGGTACCTTGGTCTTTTACCATAACCTTTCGATTTATGATACAGTACCTTCTAAGTG 2100
QY 2101 CTTTGAATTTGAACTCTACAGTATGACAGTACTATTTACATATTTGGTATCTCTCGA 2160
Db 2101 CTTTGAATTTGAACTCTACAGTATGACAGTACTATTTACATATTTGGTATCTCTCGA 2160
QY 2161 ATTCCTTTTACGATGTTGATGTCACATTTGAGTCTGCGGATGGCTCTCAAAATGGGAGA 2220
Db 2161 ATTCCTTTTACGATGTTGATGTCACATTTGAGTCTGCGGATGGCTCTCAAAATGGGAGA 2220
QY 2221 GGAAAGGATAATGGAAGAGCAGAGAAAGGCGGTAGTAGTAAAAAACAAGAAAAA 2280
Db 2221 GGAAAGGATAATGGAAGAGCAGAGAAAGGCGGTAGTAGTAAAAAACAAGAAAAA 2280
QY 2281 GAAAGTTCCGCTTGTAGCCGAGAGATCAAAATGAGCCAGCATATCAGAACATGTGTGC 2340
Db 2281 GAAAGTTCCGCTTGTAGCCGAGAGATCAAAATGAGCCAGCATATCAGAACATGTGTGC 2340

```

; APPLICANT: MARQUIS, Joseph P.
; FILE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT
; TITLE REFERENCE: PF-0899 USN
; CURRENT APPLICATION NUMBER: US/10/467,433
; CURRENT FILING DATE: 2003-08-06
; PRIOR APPLICATION NUMBER: PCT/US02/03709
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/268,117
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: US 60/269,618
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: US 60/271,118
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: US 60/274,486
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 60/274,436
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/334,229
; PRIOR FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: US 60/353,284
; PRIOR FILING DATE: 2002-02-01
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PERL Program
; SEQ ID NO 23
; LENGTH: 2488
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 7111920CBI
; US-10-467-433-23

```

RESULT 2	
US-10-467-433-23	
; Sequence 23, Application US/10467433	
; Publication No. US20040087773A1	
; GENERAL INFORMATION:	
; APPLICANT: INCYTE CORPORATION;	
/	APPLICANT: LAL, Preeti G.; BAUGHN, Mariah R.;
/	APPLICANT: YAO, Monique G.; CHAMLA, Narinder K.;
/	APPLICANT: ELLIOTT, Vicki S.; XU, Yuming;
/	APPLICANT: HONCHELL, Cynthia D.; YUE, Henry;
/	APPLICANT: DING, Li; GIETZEN, Kimberly J.;
/	APPLICANT: ISON, Craig H.; LU, Dyung Aina M.;
/	APPLICANT: HAFALIA, April J.A.; GANDHI, Ameena R.;
/	APPLICANT: THANGARELU, Kavitha; SANJANWALA, Madhusudan M.;
/	TANG, Y. Tom; RAMKUMAR, Javalaxmi;
/	APPLICANT: GRIFFIN, Jennifer A.; SWARNAKAR, Anita;
/	APPLICANT: AZIMZAI, Yalda; SAPPERSTEIN, Stephanie K.;
/	APPLICANT: BURFORD, Neil; LEE, Ernestine A.;
/	APPLICANT: LU, Yan; TRAN, Uyen K.;

Db 541 CTTGAAATCTGTGACATTCAGAGGGAAGTAATAAAGCTGCTGTTTTGAAGAGA 600
QY 1074 AGATTTTCAGTCAATGACCTATGATTTAAATAGGCTTAACAGTGTGACAGATCTTCAGT 1133
Db 601 AGATTTTCAGTCAATGACCTATGATTTAAATAGGCTTAACAGTGTGACAGATCTTCAGT 660
QY 1134 TACAGGATGCTTAAAGATGTGAGGATGACATGCAAGAGAGTAAAGAGTACTCGAAG 1193
Db 661 TACAGGATGCTTAAAGATGTGAGGATGACATGCAAGAGAGTAAAGAGTACTCGAAG 720
QY 1194 TCGACAGAGAGAGAGATCCAGAGTTGAACTAGAACCAACAATGTTTAGCAGT 1253
Db 721 TCGACAGAGAGAGAGATCCAGAGTTGAACTAGAACCAACAATGTTTAGCAGT 780
QY 1254 ATTACGACAGTGAATTTACTCGTGTGTTACTGACAGTCTTATAGCCTTTACTAAGAA 1313
Db 781 ATTACGACAGTGAATTTACTCGTGTGTTACTGACAGTCTTATAGCCTTTACTAAGAA 840
QY 1314 AGAGACAGTGTGTTGAGAGCTCAAAAATTGATGTTCAAGCAGCAGATCTTCTTTC 1373
Db 841 AGAGACAGTGTGTTGAGAGCTCAAAAATTGATGTTCAAGCAGCAGATCTTCTTTC 900
QY 1374 TSCCATTCATATTCATTCGATCATGGCATCCAGGCCGAGATGATACAAAAGGAGA 1433
Db 901 TGCCATTCATATTCATTCGATCATGGCATCCAGGCCGAGATGATACAAAAGGAGA 960
QY 1434 TCATCCAAATATGATGGGTTTTGAACCCCTGTGACAGAGGCTTCCACCTACCTT 1493
Db 961 TCATCCAAATATGATGGGTTTTGAACCCCTGTGACAGAGGCTTCCACCTACCTT 1020
QY 1494 CCCTCGATGCAAAAATAATTAAGAGGAGAAATGTTGAACCTATTTTGCAGATTAAT 1553
Db 1021 CCCTCGATGCAAAAATAATTAAGAGGAGAAATGTTGAACCTATTTTGCAGATTAAT 1080
QY 1554 AGATAGATATAAATCTGTGAGGTTGTGAATTTAACAAATTTACATGTTATCTCGA 1613
Db 1081 AGATAGATATAAATCTGTGAGGTTGTGAATTTAACAAATTTACATGTTATCTCGA 1140
QY 1614 TTTTCTCTGTAATTTAGTGAACAGTCAACATGTTCTTCAAGATCTCTGTACAAAC 1673
Db 1141 TTTTCTCTGTAATTTAGTGAACAGTCAACATGTTCTTCAAGATCTCTGTACAAAC 1200
QY 1674 CACTTCTCGTGGATACAAAAGGCTTTTGGAACTCATCTCATGCAAGACATGGTCAA 1733
Db 1201 CACTTCTCGTGGATACAAAAGGCTTTTGGAACTCATCTCATGCAAGACATGGTCAA 1260
QY 1734 AGATGCACTTCGGTCTTTTGTGAGTCTCCGAGTGTCTTCCCAAGTGTACCTATAT 1793
Db 1261 AGATGCACTTCGGTCTTTTGTGAGTCTCCGAGTGTCTTCCCAAGTGTACCTATAT 1318
QY 1794 AATAATCACAGGCTTAAGAGCTGTATCGACTCTTGTGTTACTCTGTTGTTGCGGCATTC 1853
Db 1319 AATAATCACAGGCTTAAGAGCTGTATCGACTCTTGTGTTACTCTGTTGTTGCGGCATTC 1378
QY 1854 TGTAGTCTTATTCAGATCCATGACATTAACAGGCTCGACAGAGAGATAAGCTTGGTAT 1913
Db 1379 TGTAGTCTTATTCAGATCCATGACATTAACAGGCTCGACAGAGAGATAAGCTTGGTAT 1438
QY 1914 ATTCCTGAGGAATTTGCCACCTTCAGAGTGTATGACATTTTATTTAAATAGGAG 1973
Db 1439 ATTCCTGAGGAATTTGCCACCTTCAGAGTGTATGACATTTTATTTAAATAGGAG 1475
QY 1974 AGAAGTTGATGACGCTTCAACCATGCTGTTGAAACAGGACCCCAAGGCAACATT 2033
Db 1476 AGAAGTTGATGACGCTTCAACCATGCTGTTGAAACAGGACCCCAAGGCAACATT 1535
QY 2034 TGGCTGTTAGGTACCTGGTCTTTACATTAACCTTGGATATGATACAGTACCTTC 2093
Db 1536 TGGCTGTTAGGTACCTGGTCTTTACATTAACCTTGGATATGATACAGTACCTTC 1595
QY 2094 TAAGTGGCTTTGAATTTGAACCTTACAGTATGACAGTACTATATATATTTGGTATC 2153

Db 1596 TAAGTGGCTTTGAATTTGAACCTCTACAGTATGACAGTACTATTACATATTTGGTATC 1655
QY 2154 TCTCTGAATTCCTTTTACGCATGTTGATGTCAACATTTAGTCTGCGCATGGCTCTCAAA 2213
Db 1656 TCTCTGAATTCCTTTTACGCATGTTGATGTCAACATTTAGTCTGCGCATGGCTCTCAAA 1715
QY 2214 TGCAGAGAGAGAGATTAATGGAAGAGCAGGAGAGAGCCCTAGTAGTAAAAAAACAAGA 2273
Db 1716 TGCAGAGAGAGAGATTAATGGAAGAGCAGGAGAGAGCCCTAGTAGTAAAAAAACAAGA 1775
QY 2274 AAAAAAGAGAGTTCCGCCATTTGAGCCGAGAGATCAATAGCCAGCATATCAGACA 2333
Db 1776 AAAAAAGAGAGTTCCGCCATTTGAGCCGAGAGATCAATAGCCAGCATATCAGACA 1835
QY 2334 TGTGCTCTGGAATGTTTAAAAACAATGTTAGCATTTGACATGGAAGGACGTTGCTTCAAT 2393
Db 1836 TGTGCTCTGGAATGTTTAAAAACAATGTTAGCATTTGACATGGAAGGACGTTGCTTCAAT 1895
QY 2394 CGAAGTTTCAGCTTGTATAGTGAACAAGTTCCGTTGAGTGAACACAGGTTGCTTCAATCAACA 2453
Db 1896 CGAAGTTTCAGCTTGTATAGTGAACAAGTTCCGTTGAGTGAACACAGGTTGCTTCAATCAACA 1955
QY 2454 GTGTGATGACCCCGCCGCGCAGTGCACTACTTACAGTTCAAGGAAATGTTGACCTCAATA 2513
Db 1956 GTGTGATGACCCCGCCGCGCAGTGCACTACTTACAGTTCAAGGAAATGTTGACCTCAATA 2015
QY 2514 AATATAGCCCTCTCTCTCAGTCTCTGAACTGTATGTGCGCAGCTAGTAGCACTTTCAAC 2573
Db 2016 AATATAGCCCTCTCTCTCAGTCTCTGAACTGTATGTGCGCAGCTAGTAGCACTTTCAAC 2075
QY 2574 AGGCAAAAATGATATTTGAAATATTTGAAATATTTCTTAAACCCGACCATGAGTTTAA 2633
Db 2076 AGGCAAAAATGATATTTGAAATATTTGAAATATTTCTTAAACCCGACCATGAGTTTAA 2135
QY 2634 AGGTTGCCAAACCCCACTTTTGGTGTATGAAGTTATTTGCGAGAGACACAAAAGGAAT 2693
Db 2136 AGGTTGCCAAACCCCACTTTTGGTGTATGAAGTTATTTGCGAGAGACACAAAAGGAAT 2195
QY 2694 CTAAAGTTCTCTGAAATTTGATTTCTCTGCTCATATAATTTTCTGTTGAAACTTG 2753
Db 2196 CTAAAGTTCTCTGAAATTTGATTTCTCTGCTCATATAATTTTCTGTTGAAACTTG 2255
QY 2754 TTTGAGAGAGACTGGGAGGTGGCCATAAAGGGGAGAGTCTTCTTTCAGACCCAACTCT 2813
Db 2256 TTTGAGAGAGACTGGGAGGTGGCCATAAAGGGGAGAGTCTTCTTTCAGACCCAACTCT 2315
QY 2814 TAGAGGACATCACACAGGCTCCACATCACGGGAAGTGAAGTGGATTTCTTGGGTAAACA 2873
Db 2316 TAGAGGACATCACACAGGCTCCACATCACGGGAAGTGAAGTGGATTTCTTGGGTAAACA 2375
QY 2874 CTCATTATAGGAATATCTTTTAGTTTGAACGCTTATATGATGATGAATGAATGAATGCTGT 2933
Db 2376 CTCATTATAGGAATATCTTTTAGTTTGAACGCTTATATGATGATGAATGAATGAATGCTGT 2435
QY 2934 TTTAAAGTGGTTTATTTTCCATGGAAGAACTGGTCTTATTGAAT 2981
Db 2436 TTTAAAGTGGTTTATTTTCCATGGAAGAACTGGTCTTATTGAAT 2483

RESULT 3

US-09-814-353-20725/c
; Sequence 20725, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21

```
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20725
; LENGTH: 2610
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1, 2, 3, 4
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-20725

Query Match          72.0%; Score 2228.4; DB 10; Length 2610;
Best Local Similarity 97.9%; Pred. No. 0;
Matches 2299; Conservative 0; Mismatches 21; Indels 28; Gaps 3;

QY 426 CGCATGCTGCAGCTCCGGCTCGGGCTGGGCTGAGAGGGAGGGGGCGGCGGCCGA 485
DB 2610 CGCATGCTGCAGCTCCGGCTCGGGCTGGGCTGAGAGGGAGGGGGCGGCGGCCGA 2551

QY 486 GCGGGCTGCTTTATTTCCGTGCTCCGACAGTGGTGGGCGGCGGCTGACCCAGGAGA 545
DB 2550 GCGGGCTGCTTTATTTCCGTGCTCCGACAGTGGTGGGCGGCGGCTGACCCAGGAGA 2491

QY 546 AGTAGGCATTAATGTTATGAAGCTCTCTGTAGATGATGACGATTTCAGGATGGGAGCTCAG 605
DB 2490 AGTAGGCATTAATGTTATGAAGCTCTCTGTAGATGATGACGATTTCAGGATGGGAGCTCAG 2431

QY 606 TATGCCAGAAAAAATGAGAAAAAGCAATACAAACTGGGTGGACATTACCCAAAGATTTGA 665
DB 2430 TATGCCAGAAAAAATGAGAAAAAGCAATACAAACTGGGTGGACATTACCCAAAGATTTGA 2371

QY 666 AGAAGCTTTCGAGAAATTAAGTTGGGAGAACTACTTCAATGATAGCTATTTGGTCTTTT 725
DB 2370 AGAAGCTTTCGAGAAATTAAGTTGGGAGAACTACTTCAATGATAGCTATTTGGTCTTTT 2311

QY 726 TGAAGCCATGCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATTGGAAA 785
DB 2310 TGAAGCCATGCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATTGGAAA 2251

QY 786 CCAAGTTTAATCGAAAAGTTCTCAATTTTGAACAGCTATCAAGGATGGCACTAFTAAAAAT 845
DB 2250 CCAAGTTTAATCGAAAAGTTCTCAATTTTGAACAGCTATCAAGGATGGCACTAFTAAAAAT 2191

QY 846 TAAAGATCTCACCTTGCTGCACTGAAGTATAGGATATGATGATACATGTTTGTGCTTTGAT 905
DB 2190 TAAAGATCTCACCTTGCTGCACTGAAGTATAGGATATGATGATACATGTTTGTGCTTTGAT 2131

QY 906 AACGTGTTAGAGGCCATTCACTGGCACAGACAGTATTTACGTGCTTTACATTCATAA 965
DB 2130 AACGTGTTAGAGGCCATTCACTGGCACAGACAGTATTTACGTGCTTTACATTCATAA 2071

QY 966 TCCAGACTTTTATAGAAGATCCTGCTATGAAGGCTTTTGTCTCTGGGAATCTTTGAAAACTG 1025
DB 2070 TCCAGACTTTTATAGAAGATCCTGCTATGAAGGCTTTTGTCTCTGGGAATCTTTGAAAACTG 2011

QY 1026 TGACATTTGCAAGGAAAAAGTAATAAGCTGCTGTTTTTGAAGAGGAGAAATTTTCAGTC 1085
DB 2010 TGACATTTGCAAGGAAAAAGTAATAAGCTGCTGTTTTTGAAGAGGAGAAATTTTCAGTC 1951

QY 1086 AATGACTTATGATTTAAATGGCTAAACAGTGTGACAGATCTTCAGATTACAGCATGCT 1145
```

```
DB 1950 AATGACTTATGATTTAAATGCTAACAGTGTGACAGATCTTCGAGTTACAGCATGCT 1891
QY 1146 AAAAGATGTGAGGATGACATGCAAAAGAGAGTAAAGAGTACTCGAAGTCGAAAGAGA 1205
DB 1890 AAAAGATGTGAGGATGACATGCAAAAGAGAGTAAAGAGTACTCGAAGTCGAAAGAGA 1831
QY 1206 AGAAAGAGATCCAGAGTTGAACTAGAAACCAACAAATGTTTACAGCAGTATTCAGCAGT 1265
DB 1830 AGAAAGAGATCCAGAGTTGAACTAGAAACCAACAAATGTTTACAGCAGTATTCAGCAGT 1771
QY 1266 GAAATTTACTCGTGTGTTACTGACAGTCTTATAGCCTTTACTAAAGAAAGAGACAGTGC 1325
DB 1770 GAAATTTACTCGTGTGTTACTGACAGTCTTATAGCCTTTACTAAAGAAAGAGACAGTGC 1711
QY 1326 TGTTCAGAAAGCTCAAAAATTTGATGTTCAAGCAGCAGATCTTCTTTCTGCCATTCATAA 1385
DB 1710 TGTTCAGAAAGCTCAAAAATTTGATGTTCAAGCAGCAGATCTTCTTTCTGCCATTCATAA 1651
QY 1386 TTCAATTGCATCATGGCATTCCAGGCCCGAGAAATGATATACAAAAGAGAGATCATCCAATAT 1445
DB 1650 TTCAATTGCATCATGGCATTCCAGGCCCGAGAAATGATATACAAAAGAGAGATCATCCAATAT 1591
QY 1446 GATGGGTTTTGAACCCCTTCGTAACACAGAGGCTACTTCCACCTACCTTCCCTCGATATGC 1505
DB 1590 GATGGGTTTTGAACCCCTTCGTAACACAGAGGCTACTTCCACCTACCTTCCCTCGATATGC 1531
QY 1506 AAAAAATAATTTAAAGGGAAGAAATGGTGAACATAATTTTGCAGAGTAAATAGATAGAAATAA 1565
DB 1530 AAAAAATAATTTAAAGGGAAGAAATGGTGAACATAATTTTGCAGAGTAAATAGATAGAAATAA 1471
QY 1566 AACTGCTGTGAGGTTGTGAATTTTAAACAAATTTTACATTTGATCTCTGATTTTCTGTGA 1625
DB 1470 AACTGCTGTGAGGTTGTGAATTTTAAACAAATTTTACATTTGATCTCTGATTTTCTGTGA 1411
QY 1626 ATTTAGTGAACAGTCCACATGTGTTCTTCAAGATCTCTGTATCAAAACACCTTTTCCCTGGT 1685
DB 1410 ATTTAGTGAACAGTCCACATGTGTTCTTCAAGATCTCTGTATCAAAACACCTTTTCCCTGGT 1351
QY 1686 GGATAACAAAAAGGTCTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACATTCG 1745
DB 1350 GGATAACAAAAAGGTCTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACATTCG 1291
QY 1746 GTCTTTTGTGAGATCTCCCGAGTGTCTTCCCAAGTGTCTACCTATATATATATCAGCAG 1805
DB 1290 GTCTTTTGTGAG--TCTCCGGTGTCTTCCCAAGTGTCTACCTATATATATATCAGCAG 1233
QY 1806 GCTAAGGACTGTATCGACTCCTTTGTTTACTCACTGTGTTCCGCCATTTCTGTAGTCTTATT 1865
DB 1232 GCTAAGGACTGTATCGACTCCTTTGTTTACTCACTGTGTTCCGCCATTTCTGTAGTCTTATT 1173
QY 1866 CAGATCCATGGAATACAGGGCTCGACAGAGAGATAAGCTTGGTCAATATTTCTTGAGGAA 1925
DB 1172 CAGATCCATGGAATACAGGGCTCGACAGAGAGATAAGCTTGGTCAATATTTCTTGAGGAA 1113
QY 1926 TTTGCCACCTTGCAGGATGAGTTTATGACATTTTATTTTATAGGACAGAGAGGTTGATG 1985
DB 1112 TTTGCCACCTTGCAGGATG-----AGGACAGAGAGGTTGATG 1076
QY 1986 CAGCGCTTACACATGCTGTTGAAAACAGGAAACCCAAAGGCAACATTTGGCTGTTTAG 2045
DB 1075 CAGCGCTTACACATGCTGTTGAAAACAGGAAACCCAAAGGCAACATTTGGCTGTTTAG 1016
QY 2046 GTACCTGGGTCCTTTTACCAATAACCTTTCGCATATGATACAGTACCTTCTAAGTGGCTTTG 2105
DB 1015 GTACCTGGGTCCTTTTACCAATAACCTTTCGCATATGATACAGTACCTTCTAAGTGGCTTTG 956
QY 2106 AATTGGAACTTACAGTATGACAGAGTACTTATACATATATTTGGTATCTCTCTGAATTC 2165
DB 955 AATTGGAACTTACAGTATGACAGAGTACTTATACATATATTTGGTATCTCTCTGAATTC 896
QY 2166 TTTTACGATGTTGATGTCAACATTCAGTGTGTCGATGCTCTCAATGGCAGAGGAAA 2225
```

```
Db 895 TTTAGCGATGTTGATGTCACATTTAGTCTGTCGCGATGGCTCTCAATGGCAGAGAAA 836
QY 2226 GGATTAATGGAGAGCAGAGAAAAGCCGCTAGTAGTAAATAAACAAGAAAAGAAAG 2285
Db 835 GGATTAATGGAGAGCAGAGAAAAGCCGCTAGTAGTAAATAAACAAGAAAAGAAAG 776
QY 2286 TTCGCCCAATGAGCGGAGAGATCAAAATGAGCAAGCATATCAGAACATGTTGCTGGAA 2345
Db 775 TTCGCCCAATGAGCGGAGAGATCAAAATGAGCAAGCATATCAGAACATGTTGCTGGAA 716
QY 2346 TGTTTAAACCATGCTAGCATTTGACATGCGCGCAAAAGTAGCTAAACCGAAGTTTGAGC 2405
Db 715 TGTTTAAACCATGCTAGCATTTGACATGCGCGCAAAAGTAGCTAAACCGAAGTTTGAGC 656
QY 2406 TTGATAGTGAACAAGTTCGGTATGAACAACAGGTTTGCTCCATTTCAACAGTGTGATGACCC 2465
Db 655 TTGATAGTGAACAAGTTCGGTATGAACAACAGGTTTGCTCCATTTCAACAGTGTGATGACCC 596
QY 2466 CGCGCCAGTGCACTACTTACAGTTTCAAGGAAATGCTGACCTCAATAAATATAGCCCTC 2525
Db 595 CGCGCCAGTGCACTACTTACAGTTTCAAGGAAATGCTGACCTCAATAAATATAGCCCTC 536
QY 2526 CTCCTCAGTCTCCTGAACTGATGTCGAGCTAGTAAGCACTTTCAACAGGCAAAAATGA 2585
Db 535 CTCCTCAGTCTCCTGAACTGATGTCGAGCTAGTAAGCACTTTCAACAGGCAAAAATGA 476
QY 2586 TATTGGAAAATATTCCTAACCGGACCAATGAGTTAATAGAAATTTAAAGTTGCAAAAC 2645
Db 475 TATTGGAAAATATTCCTAACCGGACCAATGAGTTAATAGAAATTTAAAGTTGCAAAAC 416
QY 2646 CCAACTTTGCTGATGAAGTTATTGGCAGGAGGACACAAAAGGAATCTAAAGTTCTC 2705
Db 415 CCAACTTTGCTGATGAAGTTATTGGCAGGAGGACACAAAAGGAATCTAAAGTTCTC 356
QY 2706 CTGAATTTGATTTCTCTCTCATAAATATTTTCTGTTGTGAAACTGTTTGGAGAGAC 2765
Db 355 CTGAATTTGATTTCTCTCTCATATA---ATAATTCCTGTTGTGAAACTGTTGGAGAGACTG 299
QY 2766 TGGGGAGG 2773
Db 298 GGAGGTGG 291
```

RESULT 4

```
US-09-814-353-15927/c
; Sequence 15927, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-25
; PRIOR FILING DATE: 2000-06-15
; PRIOR FILING DATE: 2000-07-25
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15927
; LENGTH: 745
```

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 571

; OTHER INFORMATION: n = A,T,C or G

US-09-814-353-15927

```
Query Match 18.1%; Score 560.8; DB 10; Length 745;
Best Local Similarity 94.8%; Pred. No. 1.3e-142;
Matches 621; Conservative 0; Mismatches 8; Indels 26; Gaps 3;
```

```
QY 1399 GGCATCCAGGCCGAGATGATACATAAAGAGAGATCATCAATTTATCATGGGTTTTGAA 1458
Db 739 GGCATCCAGGCCGAGATGATACATAAAGAGAGATCATCAATTTATCATGGGTTTTGAA 680
QY 1459 CCCCTTGTGAACAGAGAGCTACTTCCACCTTCCCTCCGATATGCAAAAATAATTTAA 1518
Db 679 CCCCTTGTGAACAGAGAGCTACTTCCACCTTCCCTCCGATATGCAAAAATAATTTAA 620
QY 1519 AGGGAAGAAATGGTGAACATTTTGCAGAGATTAATAGATAAATAAACTGTCTGTGAG 1578
Db 619 AGGGAAGAAATGGTGAACATTTTGCAGAGATTAATAGATAAATAAACTGTCTGTGAG 560
QY 1579 GTTGTGAATTTAAACAATTTTACATTTGATCTCTGGA-TTTTTCTGTGAATTTAGTGA 1637
Db 559 GTTGTGAATTTAAACAATTTTACATTTGATCTCTGGAATTTTCTGTGAATTTAGTGA 500
QY 1638 GTCACCATGCTGCTTTTCAAGATCTCTTTACAAACCACTTTCTGTGGTGGATAACAAAA 1697
Db 499 GTCACCATGCTGCTTTTCAAGATCTCTTTACAAACCACTTTCTGTGGTGGATAACAAAA 440
QY 1698 GGTCTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACTTCGGTCTTTTGTGAG 1757
Db 439 GGTCTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACTTCGGTCTTTTGTGAG 380
QY 1758 ATCTCTCGAGTGCTTTCCCGCAAGTCTACCTATATATAATATACCCAGGCTAAGGACGT 1817
Db 379 TCCTCGG--GTGCTTTCCCGCAAGTCTACCTATATATAATATACCCAGGCTAAGGACGT 322
QY 1818 ATCGACTCTTTTGTACTCACTGTGTGGCCACTTCTGTAGTCTTATTCAGATCCATGGA 1877
Db 321 ATCGACTCTTTTGTACTCACTGTGTGGCCACTTCTGTAGTCTTATTCAGATCCATGGA 262
QY 1878 CATACAGGCTCGACAGAGATAGCTTGGTCAATTTCTTGGAGAAATTTGCCACCTTG 1937
Db 261 CATACAGGCTCGACAGAGATAGCTTGGTCAATTTCTTGGAGAAATTTGCCACCTTG 202
QY 1938 CAGGATGAGTTTATGACATTTTATTTAATAGGAGAGAGGTTGATCGAGGCTTCACA 1997
Db 201 CAGGATG-----AGGAGAGAGAGGTTGATCGAGGCTTCACA 165
QY 1998 CCATGCTGTTGAAAACAGAAACCCCAAGGCAACATTTGGCTGTTTAGTACCTG 2052
Db 164 CCATGCTGTTGAAAACAGAAACCCCAAGGCAACATTTGGCTGTTTAGTACCTG 110
```

RESULT 5

```
US-09-814-353-16155
; Sequence 16155, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
```

```
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16155
; LENGTH: 610
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-16155

Query Match      16.5%; Score 509.4; DB 10; Length 610;
Best Local Similarity 99.8%; Pred. No. 1.4e-128;
Matches 510; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1184 GTACTCGAAGTCGACAAAGGAGAGAAAGACATCCAGAAAGTTGAACTAGAACACCAACAAT 1243
Db 100 GTACTCGAAGTCGACAAAGGAGAGAAAGACATCCAGAAAGTTGAACTAGAACACCAACAAT 159
QY 1244 GTTTAGCAGATTACGACAGAGTGAATTTACTCGTGTGTACTGACAGAGTGTATAGCCT 1303
Db 160 GTTTAGCAGATTACGACAGAGTGAATTTACTCGTGTGTACTGACAGAGTGTATAGCCT 219
QY 1304 TTACTAAGAAAGACACAGTGTCTGTCAGAAGCTCAAAAATTGATGGTTCAAGCAGCAG 1363
Db 220 TTACTAAGAAAGACACAGTGTCTGTCAGAAGCTCAAAAATTGATGGTTCAAGCAGCAG 279
QY 1364 ATCTTCTTTCGCCATTAATTAATTCATGTCATCATGSCATCCAGGCCAGAAATGATACTA 1423
Db 280 ATCTTCTTTCGCCATTAATTAATTCATGTCATCATGSCATCCAGGCCAGAAATGATACTA 339
QY 1424 CAAAGGAGATCATCCAAATTAATGATGGTTTGAACCCCTGTGACACGAGGCTACTTC 1483
Db 340 CAAAGGAGATCATCCAAATTAATGATGGTTTGAACCCCTGTGACACGAGGCTACTTC 399
QY 1484 CACCTACCTTCCCTCGATATGCAAAAATAATTAAGGGAGAAATGGTGAACCTATTTCG 1543
Db 400 CACCTACCTTCCCTCGATATGCAAAAATAATTAAGGGAGAAATGGTGAACCTATTTCG 459
QY 1544 CAAGATTATAGATAGAAATAAAACTGTCTGTGAGGTGTGAAATTTAAACAAATTTACATT 1603
Db 460 CAAGATTATAGATAGAAATAAAACTGTCTGTGAGGTGTGAAATTTAAACAAATTTACATT 519
QY 1604 GTATCCTGGATTCTTCTGTGAAATTTAGTGAACAGTCAACATGTGTCTTCAAGATCTC 1663
Db 520 GTATCCTGGATTCTTCTGTGAAATTTAGTGAACAGTCAACATGTGTCTTCAAGATCTC 579
QY 1664 TGTACAAACCACTTTCCTGGTGGAATAACAA 1694
Db 580 TGTACAAACCACTTTCCTGGTGGAATAACAA 610
```

RESULT 6

```
US-09-814-353-3218/c
; Sequence 3218, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
```

```
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3218
; LENGTH: 602
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 561, 590
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-3218
```

```
Query Match      16.0%; Score 496.2; DB 10; Length 602;
Best Local Similarity 93.9%; Pred. No. 5.8e-125;
Matches 568; Conservative 0; Mismatches 10; Indels 27; Gaps 4;

QY 1450 GGTTTTGAACCCCTGTGACACGAGGCTACTTCCACCTACCTTCCCTCCGATATGC-AAA 1508
Db 602 GGTTTTGAACCCCTGTGACACGAGGCTACTTCCACCTACCTTCCCTCCGATATGC-AAA 543
QY 1509 AATPAATTAAGGAGAAATGCTGAATTTTGAAGATTAAATAGATAGATAAATAAC 1568
Db 542 AATPAATTAAGGAGAAATGCTGAATTTTGAAGATTAAATAGATAGATAAATAAC 483
QY 1569 TGTCTGTGAGGTGTGAATTTTAAATAATTTACAAATTTACATTTATCTCTGGA-TTTTTCTGTGAAT 1627
Db 482 TGTCTGTGAGGTGTGAATTTTAAATAATTTACAAATTTACATTTATCTCTGGAATTTTCTGTGAAT 423
QY 1628 TTAGTCAACAGTCACCATGTGTCTTCAAGTCTCTGTGTACAAACCACTTCTCTGGTG 1687
Db 422 TTAGTCAACAGTCACCATGTGTCTTCAAGTCTCTGTGTACAAACCACTTCTCTGGTG 363
QY 1688 ATAACAAAAGGTCCTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACTTCGGT 1747
Db 362 ATAACAAAAGGTCCTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACTTCGGT 303
QY 1748 CTTTGTGAGATCCCTCCGAGTGTCTTCCCGAAGTGTACCTATATATAATATCAACAGGC 1807
Db 302 CTTTGTGAG--TCCTCCGGTGTCTTCCCGAAGTGTACCTATATATAATATCAACAGGC 245
QY 1808 TAAGGACTGTATCGACTCTTGTGTACTCACCTGTGTTCGGCCATCTCTGTAGTCTTATTC 1867
Db 244 TAAGGACTGTATCGACTCTTGTGTACTCACCTGTGTTCGGCCATCTCTGTAGTCTTATTC 185
QY 1868 GATCCATGGACATAACAGGGCTCGACAGAGATAGCTTTGGTCAATATTTCTTGAGGAAT 1927
Db 184 GATCCATGGACATAACAGGGCTCGACAGAGATAGCTTTGGTCAATATTTCTTGAGGAAT 125
QY 1928 TGCCACCTTGAGGATGATTTATGACATTTTATTTTATAGGAGAGAGGTGTGATGCA 1987
Db 124 TGCCACCTTGAGGATG-----AGGCAGAGAGAGGTGTGATGCA 88
QY 1988 GGGCTTTCACCATCTGTGTGAACAGGAACCCCAAGGCAACATTTGCCCTGTTAGGT 2047
Db 87 GGGCTTTCACCATCTGTGTGAACAGGAACCCCAAGGCAACATTTGCCCTGTTAGGT 28
QY 2048 ACCTG 2052
Db 27 ACCTG 23
```


RESULT 7

US-09-814-353-9543/c
; Sequence 9543, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814.353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9543
; LENGTH: 602
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 561..590
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-9543

Query Match 16.0%; Score 496.2; DB 10; Length 602;

Best Local Similarity 93.9%; Pred. No. 5.8e-125;

Matches 568; Conservative 0; Mismatches 10; Indels 27; Gaps 4;

QY	1450	GGTTTGAACCCCTGTGTAACAGAGGCTACTTCCACCTACTTCCCTCGATATGC-AAA	1508
Db	602	GGTTTGAACCCCTGTGTAACAGAGGCTACTTCCACCTACTTCCCTCGATATGC-AAA	543
QY	1509	AATAATTAAGGGGAAGAAATGGTGAACATAATTTGCAAGATTAATAGATAGATAAAT	1568
Db	542	AATAATTAAGGGGAAGAAATGGTGAACATAATTTGCAAGATTAATAGATAGATAAAT	483
QY	1569	TGCTGTGAGGTTGTGAATTAACAAATTAATGATGATCTCTGGA-TTTTTCTGTGAAT	1627
Db	482	TGCTGTGAGGTTGTGAATTAACAAATTAATGATGATCTCTGGAATTTTTCTGTGAAT	423
QY	1628	TTAGTGAACAGTCACCATGTGTCTTTTCAAGATCTCTGTATCAAAACCACTTTCCTGGTGG	1687
Db	422	TTAGTGAACAGTCACCATGTGTCTTTTCAAGATCTCTGTATCAAAACCACTTTCCTGGTGG	363
QY	1688	ATAACAAAAGGTCTTTTGAACATCATCTCATGCAAGACATGTTGAAAGATGCACCTTCGT	1747
Db	362	ATAACAAAAGGTCTTTTGAACATCATCTCATGCAAGACATGTTGAAAGATGCACCTTCGT	303
QY	1748	CTTTTGTGAGTCTCCGAGTGTCTTTCCCAAGTGTACCTATATATATATATATATATATAT	1807
Db	302	CTTTTGTGAGTCTCCGAGTGTCTTTCCCAAGTGTACCTATATATATATATATATATATAT	245
QY	1808	TAAGGACTGTATCGACTCTTTTGTACTCATCTGTGTGCGGCATCTGTAGTCTTATTC	1867
Db	244	TAAGGACTGTATCGACTCTTTTGTACTCATCTGTGTGCGGCATCTGTAGTCTTATTC	185
QY	1868	GATCCATGACATTAACAGGCTCGACAGAGATAGCTTGTGATATCTTGTGAGGAAT	1927
Db	184	GATCCATGACATTAACAGGCTCGACAGAGATAGCTTGTGATATCTTGTGAGGAAT	125

QY	1928	TGCCACCTTCAGGATGAGTTTATGACATTTTATTTAATAGGCAGAGAGTTGATGCA	1987
Db	124	TGCCACCTTCAGGATGAGTTTATGACATTTTATTTAATAGGCAGAGAGTTGATGCA	88
QY	1988	GCCTTCACACCATGCTGTTGAAACAGAAACCCCAAGGCAACATTTGGCTCTTTAGGT	2047
Db	87	GCCTTCACACCATGCTGTTGAAACAGAAACCCCAAGGCAACATTTGGCTCTTTAGGT	28
QY	2048	ACCTG 2052	
Db	27	ACCTG 23	

RESULT 8

US-09-918-995-20914
; Sequence 20914, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 20914
; LENGTH: 482
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(482)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-20914

Query Match 14.6%; Score 451.2; DB 10; Length 482;

Best Local Similarity 99.3%; Pred. No. 1.1e-112;

Matches 453; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY	491	CGTGGTTATTTCCGTGTCGCGACAGTGCCTGGCGCGCGGTGACCCAGGAGAGTAG	550
Db	26	CGAGCTTATTTCCGTGTCGCGACAGTGCCTGGCGCGCGGTGACCCAGGAGAGTAG	85
QY	551	GCATAATGGTTATGAAGCTTCTGTAGATGATGACGATTCAGGATGGAGCTCAGTATGC	610
Db	86	GCATAATGGTTATGAAGCTTCTGTAGATGATGACGATTCAGGATGGAGCTCAGTATGC	145
QY	611	CAGAAAAAATCGAGAAAAGCAATACAACTGGTGGACATTTACCAAGATTTTGAAGAAG	670
Db	146	CAGAAAAAATCGAGAAAAGCAATACAACTGGTGGACATTTACCAAGATTTTGAAGAAG	205
QY	671	CTTGTGAGAAATTAAGTTGGGAGAACTACTTCATGATAGCTATTTGGCTCTTTTGAAG	730
Db	206	CTTGTGAGAAATTAAGTTGGGAGAACTACTTCATGATAGCTATTTGGCTCTTTTGAAG	265
QY	731	CCATGCTGTCTATTGAATGATGATCCCAAGATGGATCTGGCATGATTCGAAACCAAG	790
Db	266	CCATGCTGTCTATTGAATGATGATCCCAAGATGGATCTGGCATGATTCGAAACCAAG	325
QY	791	TTAATCGAAAAAGTTCTCAATTTTGAACAAAGCTATCAAGATGGCACTATTAATAAAG	850
Db	326	TTAATCGAAAAAGTTCTCAATTTTGAACAAAGCTATCAAGATGGCACTATTAATAAAG	385
QY	851	ATCTCACCTTGGCTGAACCTGATAGGATATGATACATGTTTGTCTGTTGATAACGT	910
Db	386	ATCTCACCTTGGCTGAACCTGATAGGATATGATACATGTTTGTCTGTTGATAACGT	445
QY	911	GGTTAGAGGCCATTCCTCTGGCACAGACAGTATTTA 946	

Db 446 GGTAGAGGCCATTCCTGTCACAGACAGTATTTA 481

RESULT 9

US-09-814-353-3455
; Sequence 3455, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John

; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B

; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: US 60/191,031

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: US 60/207,124

; PRIOR FILING DATE: 2000-05-25

; PRIOR APPLICATION NUMBER: US 60/211,940

; PRIOR FILING DATE: 2000-06-15

; PRIOR APPLICATION NUMBER: US 60/216,820

; PRIOR FILING DATE: 2000-07-07

; PRIOR APPLICATION NUMBER: US 60/220,661

; PRIOR FILING DATE: 2000-07-25

; PRIOR APPLICATION NUMBER: US 60/257,672

; PRIOR FILING DATE: 2000-12-21

; NUMBER OF SEQ ID NOS: 22037

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3455

; LENGTH: 434

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-814-353-3455

Query Match 14.0%; Score 434; DB 10; Length 434;

Best Local Similarity 100.0%; Pred. No. 5.3e-108;

Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1186	ACTCGAAGTCGACGAGAGAGAAAGAGATCCAGAGTTGAACTAGAACACCAACAATGT	1245
Db	1	ACTCGAAGTCGACGAGAGAGAGATCCAGAGTTGAACTAGAACACCAACAATGT	60
QY	1246	TTAGCAGTATTACGACAGTGAAATTTACTCGTGTGTTACTGACAGTGTCTTATAGCCTTT	1305
Db	61	TTAGCAGTATTACGACAGTGAAATTTACTCGTGTGTTACTGACAGTGTCTTATAGCCTTT	120
QY	1306	ACTAAGAAAGACACCAAGTCTGTTGCAGAGCTCAAAAATTGATGTTCAAGCAGCAGAT	1365
Db	121	ACTAAGAAAGACACCAAGTCTGTTGCAGAGCTCAAAAATTGATGTTCAAGCAGCAGAT	180
QY	1366	CTTCTTTCTGCCATTCATTAATTCATTCATGCGATCCAGGCCAGAGATGATCTACA	1425
Db	181	CTTCTTTCTGCCATTCATTAATTCATTCATGCGATCCAGGCCAGAGATGATCTACA	240
QY	1426	AAAGGAGATCATCCAAATTATGATGGGTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCA	1485
Db	241	AAAGGAGATCATCCAAATTATGATGGGTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCA	300
QY	1486	CCTACCTTCCCTCGATGCAAAAATAATTAAAGGGAAGAAATGTTGAATTTTGA	1545
Db	301	CCTACCTTCCCTCGATGCAAAAATAATTAAAGGGAAGAAATGTTGAATTTTGA	360
QY	1546	AGATTAAATAGATAGATAAAAATCTGCTGTGAGGTTGTGAATTTAAACAAATTTACATTTG	1605
Db	361	AGATTAAATAGATAGATAAAAATCTGCTGTGAGGTTGTGAATTTAAACAAATTTACATTTG	420
QY	1606	ATCCTGGATTTTTT 1619	
Db	421	ATCCTGGATTTTTT 434	

RESULT 11

RESULT 10

US-09-814-353-9771
; Sequence 9771, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:

; APPLICANT: Lee, John

; APPLICANT: Thompson, Pamela

; APPLICANT: Lillie, James

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B

; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: US 60/191,031

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: US 60/207,124

; PRIOR FILING DATE: 2000-05-25

; PRIOR APPLICATION NUMBER: US 60/211,940

; PRIOR FILING DATE: 2000-06-15

; PRIOR APPLICATION NUMBER: US 60/216,820

; PRIOR FILING DATE: 2000-07-07

; PRIOR APPLICATION NUMBER: US 60/220,661

; PRIOR FILING DATE: 2000-07-25

; PRIOR APPLICATION NUMBER: US 60/257,672

; PRIOR FILING DATE: 2000-12-21

; NUMBER OF SEQ ID NOS: 22037

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 9771

; LENGTH: 434

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-814-353-9771

Query Match 14.0%; Score 434; DB 10; Length 434;

Best Local Similarity 100.0%; Pred. No. 5.3e-108;

Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1186	ACTCGAAGTCGACGAGAGAGAAAGAGATCCAGAGTTGAACTAGAACACCAACAATGT	1245
Db	1	ACTCGAAGTCGACGAGAGAGAGATCCAGAGTTGAACTAGAACACCAACAATGT	60
QY	1246	TTAGCAGTATTACGACAGTGAAATTTACTCGTGTGTTACTGACAGTGTCTTATAGCCTTT	1305
Db	61	TTAGCAGTATTACGACAGTGAAATTTACTCGTGTGTTACTGACAGTGTCTTATAGCCTTT	120
QY	1306	ACTAAGAAAGACACCAAGTCTGTTGCAGAGCTCAAAAATTGATGTTCAAGCAGCAGAT	1365
Db	121	ACTAAGAAAGACACCAAGTCTGTTGCAGAGCTCAAAAATTGATGTTCAAGCAGCAGAT	180
QY	1366	CTTCTTTCTGCCATTCATTAATTCATTCATGCGATCCAGGCCAGAGATGATCTACA	1425
Db	181	CTTCTTTCTGCCATTCATTAATTCATTCATGCGATCCAGGCCAGAGATGATCTACA	240
QY	1426	AAAGGAGATCATCCAAATTATGATGGGTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCA	1485
Db	241	AAAGGAGATCATCCAAATTATGATGGGTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCA	300
QY	1486	CCTACCTTCCCTCGATGCAAAAATAATTAAAGGGAAGAAATGTTGAATTTTGA	1545
Db	301	CCTACCTTCCCTCGATGCAAAAATAATTAAAGGGAAGAAATGTTGAATTTTGA	360
QY	1546	AGATTAAATAGATAGATAAAAATCTGCTGTGAGGTTGTGAATTTAAACAAATTTACATTTG	1605
Db	361	AGATTAAATAGATAGATAAAAATCTGCTGTGAGGTTGTGAATTTAAACAAATTTACATTTG	420
QY	1606	ATCCTGGATTTTTT 1619	
Db	421	ATCCTGGATTTTTT 434	

US-10-242-535A-43407

```
Query Match          9.8%; Score 302.6; DB 16; Length 345;
Best Local Similarity 97.3%; Pred. No. 5.2e-72;
Matches 329; Conservative 0; Mismatches 5; Indels 4; Gaps 2;

QY 2747 AAACCTGTTTGGAGAGACTGGGAGGTGGCCATAAAGGGGCGAGAGTCTTCTTCAGACC 2806
Db 1 AACCTGTTTGGAGAGACTGGGAGGTGGCCCTAAAGGGGCGAGAGTCTTCTTCAGACC 60

QY 2807 CAACCTTTAGAGGCGACATCACCCAGGCTCCACATCAGCGGAAGTGCAGATGATTTCTTGG 2866
Db 61 CAACCTTTAGAGGCGACATCACCCAGGCTCCACATCAGCGGAAGTGCAGATGATTTCTTGG 120

QY 2867 GTAACTCACTATTATAGGATACCTTTTAGTTGACAGCCTTATATGACATGAATGAAAA 2926
Db 121 GTAACTCACTATTATAGGATACCTTTTAGTTGACAGCCTTATATGACATGAATGAAAA 180

QY 2927 CTGCTGTTTTAAAGTGGTTTATTATGTTCCATGGAAGAACTGGTCTTATTGAATGCATT 2986
Db 181 CTGCTGTTTTAAAGTGGTTTATTATGTTCCATGGAAGAACTGGTCTTATTGAATGCATT 240

QY 2987 GATGAACGTTATAT- -GGTTTTATTACAGATTTTAATCACAATCATTTTTTATGAATGAT 3044
Db 241 GATGAACGTTATATGTTTATTATGTTTATTACAGATTTTAATCACAATCATTTTTTATGAATGAT 300

QY 3045 TGAGTGAATAATAGTGGTTTATAAA- -GGTTAATAAATTT 3080
Db 301 TGAGTGAATAATAGTGGTTTATAAAAGGGTTAATAAATTT 338
```

RESULT 14

```
US-09-783-590-5693
Sequence 5693, Application US/09783590
Patent No. US20020110850A1
```

GENERAL INFORMATION:

```
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5693
```

LENGTH: 501

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (226)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (289)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (315)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (353)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (381)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (383)
```

```
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (390)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (403)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (416)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (425)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (428)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (440)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (449)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (455)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (466)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (479)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (487)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-783-590-5693
```

```
Query Match          8.7%; Score 268.6; DB 9; Length 501;
Best Local Similarity 93.2%; Pred. No. 1.5e-62;
Matches 313; Conservative 0; Mismatches 17; Indels 6; Gaps 3;

QY 2223 AAAGGATAATGGAAGAGCAGCAGAAAGCGGTAGTAGTAAATAAACAAGAAAAAAGA 2282
Db 7 AAAGGATAATGGAAGAGCAGCAGAAAGCGGTAGTAGTAAATAAACAAGAAAAAAGA 66

QY 2283 AAGTTCGCCCATTTAGCCGAGAGATCACAAATGAGCCAGCATATCAGAAATGTGCTG 2342
Db 67 AAGTTCGCCCATTTAGCCGAGAGATCACAAATGAGCCAGCATATCAGAAATGTGCTG 126

QY 2343 GAATGTTTAAACCATTTGAGCATTTGACATGAGCGGCAAGTACGTAACCGAAGTTTG 2402
Db 127 GAATGTTTAAACCATTTGAGCATTTGACATGAGCGGCAAGTACGTAACCGAAGTTTG 186

QY 2403 AGCTT-GATAGTGAACAAGTTCGGTATGAACACAGGTTTGCTCCATTCAACAGTGTGATG 2461
Db 187 AGCTTGGTTAGTGAACAAGTTCGGTATGAACACAGGTTTNCCTCAATCAACAGTGTGATG 246

QY 2462 ACCCGCCGCCAGTGCATCTACTTACAGTTT-CAAGGAAATGTCTGACCTCAATAAATA 2519
Db 247 ACCCGCCGCCAGTGCATCTACTTACAGTTTCAAGGAAATGTCTGACCTCAATAAATA 306

QY 2520 GCCTTCCTCC---TCAGTCTCTGAACTGTATGTGG 2552
Db 307 GCCTTCCTTCCTTCAGTTCCTGAACTGTATGTGG 342
```

RESULT 15

```
US-10-027-632-323193/c
Sequence 323193, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
```

```
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 323193
; LENGTH: 489
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-323193
```

```
Query Match          5.9%; Score 182; DB 13; Length 489;
Best Local Similarity 100.0%; Pred. No. 8.8e-39;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1431 AGATCATCCAATTATGATGGGTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACCTAC 1490
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1491 CTTCCTCGATATGCAAAATAATTTAAAGGGAAGAAATGGTGAACCTATTTTGAAGATT 1550
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1551 AATAGATAGATAAAACTGTCTGAGGTTGTGAAATTTAACAATTTACATTTGATCCT 1610
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1611 GG 1612
Db ||
Qy 221 GG 220
Db ||
```

Search completed: August 11, 2004, 13:00:44
Job time : 1439.92 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 10, 2004, 17:10:22 ; Search time 225.025 Seconds
(without alignments)
7635.278 Million cell updates/sec

Title: US-10-001-857-42

Perfect score: 3096

Sequence: 1 ttcttcacgaactccagg.....attcttgacaaaaaaaaa 3096

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	69.4	2.2	7218	1	US-08-232-463-14
C 2	56.4	1.8	320	3	US-09-165-264-11
C 3	55.2	1.8	320	3	US-09-165-264-14
C 4	54.8	1.8	320	3	US-09-165-264-7
C 5	54.8	1.8	320	3	US-09-165-264-13
C 6	54.2	1.8	318	3	US-09-165-264-12
C 7	54.2	1.8	319	3	US-09-165-264-8
C 8	54.2	1.8	4403765	3	US-09-103-840A-2
C 9	54.2	1.8	4411529	3	US-09-103-840A-1
C 10	53.2	1.7	2561	4	US-09-616-289-48
C 11	51	1.6	152331	3	US-09-128-155-16
C 12	50	1.6	1614	4	US-09-616-289-45
C 13	50	1.6	12425	4	US-09-616-289-50
C 14	48.4	1.6	23673	4	US-09-773-816-1
C 15	47.4	1.5	4403765	3	US-09-103-840A-2
C 16	47.4	1.5	4411529	3	US-09-103-840A-1
C 17	46.4	1.5	483	4	US-09-543-681A-4138
C 18	46.4	1.5	516	4	US-09-543-681A-4140
C 19	46.4	1.5	519	4	US-09-543-681A-4145
C 20	46.4	1.5	608	4	US-09-543-681A-4115
C 21	46	1.5	4638	4	US-09-023-655-1215
C 22	45.6	1.5	8961	4	US-10-204-708-80
C 23	45.2	1.5	1674	4	US-09-894-844-14
C 24	45.2	1.5	1674	4	US-09-894-844-15
C 25	45	1.5	1561	4	US-08-765-907A-11
C 26	44.8	1.4	2130	4	US-09-909-962A-6
C 27	44.8	1.4	2130	4	US-09-909-962A-7

C 28 44 1.4 832 4 US-09-521-976-2813 Sequence 2813, Ap
C 29 44 1.4 3572 4 US-09-575-574-3 Sequence 3, Appli
C 30 44 1.4 12001 1 US-08-458-568A-11 Sequence 11, Appl
C 31 43.8 1.4 832 4 US-09-621-976-2813 Sequence 2813, Ap
C 32 43.6 1.4 239 4 US-09-056-556-181 Sequence 181, App
C 33 43.6 1.4 239 4 US-09-072-596-176 Sequence 176, App
C 34 43.6 1.4 239 4 US-09-072-967-181 Sequence 181, App
C 35 43.6 1.4 2335 3 US-09-387-574-9 Sequence 9, Appli
C 36 43.6 1.4 2335 4 US-09-668-096-9 Sequence 9, Appli
C 37 43.4 1.4 1070 4 US-09-470-443-7 Sequence 7, Appli
C 38 43.4 1.4 1294 3 US-09-312-038-4 Sequence 4, Appli
C 39 43.4 1.4 1294 4 US-09-850-964-4 Sequence 4, Appli
C 40 43.4 1.4 2289 3 US-09-312-038-3 Sequence 3, Appli
C 41 43.4 1.4 2289 4 US-09-850-964-3 Sequence 3, Appli
C 42 43.4 1.4 5482 4 US-09-470-443-3 Sequence 3, Appli
C 43 43.2 1.4 1215 4 US-09-091-952A-29 Sequence 29, Appl
C 44 42.8 1.4 339 4 US-09-543-681A-4147 Sequence 4147, Ap
C 45 42.4 1.4 289 3 US-09-007-005-17 Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; PRIOR APPLICATION DATA:
; CLASSIFICATION: 435
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptzgt-Fls
; US-08-232-463-14

Query Match 2.2%; Score 69.4; DB 1; Length 7218;

[illegible][illegible]

```

; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-11

Query March 1.8%; Score 56.4; DB 3; Length 320;
Best Local Similarity 50.8%; Pred. No. 7.5e-06;
Matches 135; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

Qy 303 GCGAGAAAGAGCGGCGGCATCGGCCAAACGAGCGCGTGCAGGAGAGGGCGTGTG 362
Db 52 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 111

Qy 363 GCCGGGAGCGCGAAGTCCCCCGGAGGTAAAGGAGAGGGGGCGGGGTCCCGGGC 422
Db 112 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 171

Qy 423 ATACGCATGCGTGACAGTCCGCTCGGGCTGGGCTGNAGAGGGAGGGGGCGGGCGGC 482
Db 172 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 231

Ov 483 CGAGGCGGCGTCGTATTATTCCTGGTCCGACAGTGTGTCGCGCGGGGTGACACGGG 542

```

```

RESULT 4
US-09-165-264-7
; Sequence 7, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuraiayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-7

```

	Query Match	1.8%;	Score 54.8;	DB 3;	Length 320;
	Best Local Similarity	51.2%;	Pred. No. 2.1e-05;		
	Matches 128;	Conservative 0;	Mismatches 122;	Indels 0;	Gaps 0;
QY	303 GCGAGAGAAGAGGCGGCCATCGGCCTCGAACCAGAGCCGTGGCGCAGAGGAGGGGCTGTG	362			
Db	55 GG	114			
QY	363 GCCGGGAGCGCAGACTCCCCGGGAGATTAGGAGAGGGGGCGGGTCCGCGTCCC GGCC	422			
Db	115 GG	174			
QY	423 ATACGCATGTGGTGCA CGTGCCGTCCGGCTCTAGAGGGGAGGGGGCGGGCGGC	482			
Db	175 GG	234			
QY	483 CGAGCGCGCTGTTATTCGTGGTCCGAGACAGTGGTCGGCGCGCGGGTAC CACGG	542			
Db	235 GG	294			
QY	543 AGAAGTAGGC	552			
Db	295 GGGGGAGGC	304			

```

RESULT 5
US-09-165-264-13
; Sequence 13, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thiraiayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Primer sequence
US-09-165-264-13

```

RESULT 6
US-09-165-264-12
; Sequence 12, Application US/09165264

```

; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuraiyiah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 318
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-12

Query Match          1.8%; Score 54.2; DB 3; Length 318;
Best Local Similarity 50.2%; Pred.No.3.1e-05;
Matches 134; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

Qy      303 GCAGAGAGACAGCGCGGCCCATCGGCCAAACGAGGCGGTGGCGAGGAGAGAGGGGTGTG 362
Db      49  GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 108

Qy      363 GCCGGGGAGCGCAAGTCCCGGGAGTAAGGAGAGAGGGGCGGGGGTGC CGCTCCCGGGC 422
Db      109 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 168

Qy      423 ATACGATCGGTGCACGCTCCGCTCGGGGTGGGCTGAGAGGGGAGGGGGCGGGCGGC 482
Db      169 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 228

Qy      483 CGAGGGCGGCGTCTGTTATTTCCGTGGTCCGACACAGTGC GTGCGCGCGCGGGGTGACCA CGGG 542
Db      229 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 288

Qy      543 AGAAGTAGGCATATATGTTATGAAGC 569
Db      289 GGGGGGGGGGGGGTGGTGTGTGAAGC 315

```

```

RESULT 7
US-09-165-264-8
  Sequence 8, Application US/09165264
  Patent NO. 6197510
  GENERAL INFORMATION:
  APPLICANT: Vinavagamoothy, Thuraiayah
  TITLE OF INVENTION: Multi-Loci Genomic Analysis
  FILE REFERENCE: 44747
  CURRENT APPLICATION NUMBER: US/09/165,264
  CURRENT FILING DATE: 1998-10-01
  NUMBER OF SEQ ID NOS: 14
  SOFTWARE: PatentIn Ver. 2.1
  SEQ ID NO 8
  LENGTH: 319
  TYPE: DNA
  ORGANISM: Artificial Sequence
  FEATURE:
  OTHER INFORMATION: Description of Artificial Sequence: Primer sequence
US-09-165-264-8

```

[illegible]


```
QY 395 AGAGGGGCGGGTTCGGCGTCCCGGGCATACGATCGTGCACGCTGCCGCTCGGCTG 454
Db 801 GC CGCGGCGCGGCGCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 742
QY 455 GGTGAGAGGGGAGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 514
Db 741 GCGCGGCGAGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 682
QY 515 AGTGCCTGCGCGCGCGCGGTGACCAACGCGGAGAGTAGGC 552
Db 681 CCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 644

RESULT 11
US-09-128-155-16/c
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Par, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 09/128,155
; EARLIER FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match 1.6%; Score 51; DB 3; Length 152331;
Best Local Similarity 54.5%; Pred. No. 0.014;
Matches 102; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 346 GCAGGAGGGGTGTGTGGCGGGAGCGCGAAGTCCCGGGAGTAAGGAGAGGGGCGG 405
Db 22309 GGGAAATTATAGCTACTGCCCGGTTCGGGCGCGGGGGCGGGGCTGGGGGGGGGTGGGG 22250
QY 406 GGTGCGCGTCCCGGCATACGATGCGTCACTCCCGGTTCGGGCTGGGCTGAGAGG 455
Db 22249 GGGCGCGCGTTCGGGGGTGGGGGGGGGGTGGGGCGCGCGCGGGGGGGGGCGG 22190
QY 466 GAGGGGCGCGCGCGCGCGCGCGCGTCTGTTTTCGGTCCGACAGTCCGTGGCG 525
Db 22189 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 22130
QY 526 GCGCGGG 532
Db 22129 GGGGGGG 22123

RESULT 12
US-09-616-289-45/c
; Sequence 45, Application US/09616289
; Patent No. 6632923
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
```

```
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/616,289
; CURRENT FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 1614
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1614)
US-09-616-289-45

Query Match 1.6%; Score 50; DB 4; Length 1614;
Best Local Similarity 53.9%; Pred. No. 0.0013;
Matches 124; Conservative 0; Mismatches 105; Indels 1; Gaps 1;

QY 315 GCGCGCGCGCATCGCGCGAAGCGGAGCGGTGGCGAGGAGGGGGTGTGGCGGGGAGCGCG 374
Db 629 GCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 571
QY 375 GAAGTCCCGCGGAGTAAGGAGAGGGGCGGGGTCGCGGCTCCCGGCGATACGATCGGT 434
Db 570 CACCGCGCGGGGAGCGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 511
QY 435 GCACGCTCGCGCGCGCTGGCGTGAAGGAGGGGCGGGCGCGCGCGCGCGCGCGCGCG 494
Db 510 GCGGGGCGCGCGCTGCGCGGGCGCGAGGAGGGGGGGCGTGTGGCGCGCGCGCGCGCG 451
QY 495 GTTATTTCGTGTCGCGACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 544
Db 450 GCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 401

RESULT 13
US-09-289-50/c
; Sequence 50, Application US/09616289
; Patent No. 6632923
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/616,289
; CURRENT FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50
; LENGTH: 12425
; TYPE: DNA
; ORGANISM: Homo sapiens
```

US-09-616-289-50

Query Match 1.6%; Score 50; DB 4; Length 12425;
Best Local Similarity 53.9%; Pred. No. 0.0051;
Matches 124; Conservative 0; Mismatches 105; Indels 1; Gaps 1;

QY 315 GCGGCGCCCATCGCCGAAACGAGGCGGTGCGAGGAGGGGTGTGCGCGGGAGCGC 374
|||||
DB 3460 GCGGCTCCGGCGCGACGCGCGCGGGGGCGCGCGCGCGCGCGCGCGCGCGCG 3402
QY 375 GAAATCCCGGAGTAAGGAGAGAGGGCGCGGGTCCGCGCTCCCGGGCATACGATGCGT 434
DB 3401 CACCGCGGGGAGCGGCTGGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3342
QY 435 GCAGCTGCGGTCGGGTGGGCTGAGAGGAGGGGGCGCGCGCGCGCGCGCGCGCG 494
DB 3341 GCGGGGGCGCGCTGCGCGGGGCGCGAGGCGGGGGCGCTGTGGCGGGCGCGCGCG 3282
QY 495 GTTATTTCCGTGCGTCCGACAGTGTGCGCGCGCGCGGTGACACCGGAG 544
DB 3281 GCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTGCGG 3232

RESULT 14

US-09-773-816-1/c

; Sequence 1, Application US/09773816

; Patent No. 6340774

; GENERAL INFORMATION:

; APPLICANT: Stanford University

; APPLICANT: Khosla, Chaitan

; TITLE OF INVENTION: NON-STEROIDAL ESTROGEN-RECEPTOR

; TITLE OF INVENTION: ANTAGONISTS

; FILE REFERENCE: 28600-20210.00

; CURRENT APPLICATION NUMBER: US/09/773,816

; CURRENT FILING DATE: 2001-06-08

; PRIOR APPLICATION NUMBER: US 60/243,458

; PRIOR FILING DATE: 2000-10-25

; PRIOR APPLICATION NUMBER: US 60/179,305

; PRIOR FILING DATE: 2000-01-31

; NUMBER OF SEQ ID NOS: 1

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1

; LENGTH: 23673

; TYPE: DNA

; ORGANISM: Human

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)..(23673)

; OTHER INFORMATION: n = A,T,C or G

US-09-773-816-1

Query Match 1.6%; Score 48.4; DB 4; Length 23673;
Best Local Similarity 46.7%; Pred. No. 0.022; Indels 0; Gaps 0;

Matches 154; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

QY 182 CGCAGCGGGCGGGCTCTGAGACTCCGGGTCGCCCTCTTCCGGGAACCGCCACTAC 241
|||||
DB 5007 CGCATGCCCGCGCTCGACAGTTCGAGGTGAGGCGCAACGGTGGCACACCATC 4948
QY 242 CCAGAGCTCCGACAGAGGGTGAAGATATTTCCGGTCTCGCGATCGTCTTAATCT 301
DB 4947 GGGTTCTCGCGAGGTGTTCGACAAACGAAAGTTTCGCGCGCGACGCGCACACCTCC 4888
QY 302 CGCAGAGAGAGAGCGCGCGCATCGGCGCAACGAGCGGTGCGAGGGAGGGGTGT 361
DB 4897 CTCGTGACGAGATCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4828
QY 362 GCGCGGGGAGCGGAAGTCCCGGGAGTAAGGAGAGGGGGCGGGTCCGCGTCCCGGG 421
DB 4827 GCGTGAAGGGTCCGCGGTGCGCGGGTCTCCAGGCGACGCGACCCCGCGCTCGGCC 4768
QY 422 CATACGATGCGTGCACGCTGCGCGTCCGCTGAGAGGGAGGGGGCGCGCGG 481

DB 4767 GCGCGCGAAGACGACGACGCCGCGGGCGCGGGGTCTTCGCGTACCGAGCGGGGG 4708
QY 482 CCGAGGGCGGCTCGTTATTTCCGTGGTCCG 511
DB 4707 CCGGGCCCGGAACGGGATGTCCGGAACCG 4678

RESULT 15

US-09-103-840A-2/c

; Sequence 2, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 4403765

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: CDC 1551

; OTHER INFORMATION: "n" bases at various positions throughout the sequence

; OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match 1.5%; Score 47.4; DB 3; Length 4403765;
Best Local Similarity 50.7%; Pred. No. 1.3; Indels 0; Gaps 0;
Matches 114; Conservative 0; Mismatches 111;

QY 317 CCGCGCGCATCGCGCGAAGCGGCGGTGGCGAGGAGGGGTGTGGCGCGGAGCGCGA 376
DB 2800482 CCGTGGCAGTTGGCTGGCCGCGGTGACGGCGGGCGCGGTACGCTGGCGACCCCGG 2800423
QY 377 AGTCCCGGGAGTAAGGAGAGGGGGTTCGCGGTCCCGGGCATACGATGCGTGC 436
DB 2800422 CTGGGTGGGCTGGCGAGCGCGGGGCGCTCGCGCGCGCGCGGTCTCGCGCGGGGC 2800363
QY 437 ACGTGGCGGTTCGGCTGGGCTGAGAGGGAGGGGGCGCGCGCGCGCGCGCGGTTCGT 496
DB 2800362 CAATGGTCTGGCGCGCGCGCAACGACGCGGCGCGTCAAGCGCGCGCAACGCGCG 2800303
QY 497 TATTTCCGTGCTCCGACAGTGCCTGGCGCGCGGTGACCAACCG 541
DB 2800302 CAATGGCGCGCGCGCGGTGATGGCGGTAAACGG 2800258

Search completed: August 10, 2004, 21:58:26

Job time : 236.025 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 21:49:02 ; Search time 20 Seconds
(without alignments)
3525.419 Million cell u

Title: US-10-001-857-145

Perfect score:

Sequence: 1 MVMKASVDDDDSGWELSMPE.....KVPPEFDFAHKYFPVVKLV 733

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Minimum DB seq	length: 9
Maximum DB seq	length: 2000000000

Post-processing: Minimum Match 0%

Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:★

1: bird:*

```
2: pir2:*
```

3: pir3:★

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	584	15.2	695	2	B84495		hypothetical prote
2	548	14.2	814	2	T32223		hypothetical prote
3	270.5	7.0	733	2	S31288		MAK10 protein - ye
4	252.5	6.6	708	2	T33742		hypothetical prote
5	134	3.5	2429	1	SURDUA		spectrin alpha cha
6	132.5	3.4	1152	2	S20106		hypothetical prote
7	127	3.3	969	2	E71810		type III restricti
8	127	3.3	1089	2	S48244		NMD2 protein - yea
9	123.5	3.2	778	2	T16111		hypothetical prote
10	122.5	3.2	1607	2	T21982		hypothetical prote
11	122.5	3.2	2253	2	T30336		nuclear/mitotic ap
12	121.5	3.2	1603	1	VURKWS		vitellogenin vit-5
13	119	3.1	1196	2	T13057		KIAA0729 protein -
14	118.5	3.1	1029	2	T02576		hypothetical prote
15	117.5	3.1	3660	1	S02041		dystrophin, muscle
16	116	3.0	736	2	T00023		transcription fact
17	116	3.0	3147	2	T21328		hypothetical prote
18	114.5	3.0	1603	2	F89497		protein vit-5 [imp
19	114	3.0	1023	2	JC4013		major acidic nucle
20	114	3.0	1558	2	F71603		RESA-H3 antigen PP
21	114	3.0	1679	2	S48385		hypothetical prote
22	113.5	2.9	719	2	T05384		beta-adaptin homol
23	113.5	2.9	833	2	T43446		hypothetical prote
24	113.5	2.9	1957	2	T38077		hypothetical coile
25	113	2.9	2787	2	S45416		TEI1 protein - yea
26	113	2.9	3744	2	S46715		hypothetical prote
27	112.5	2.9	1642	2	T08880		NMDA receptor-bind
28	112	2.9	1130	2	T23104		hypothetical prote
29	112	2.9	1133	2	T23103		hypothetical prote

ALIGNMENTS

RESULT 1

B84495

hypothetical protein At2g11000 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C/Species: *Macropis chinensis* (Macropis chinensis)
C/Date: 02-Feb-2001 #sequence revision 02-Feb-2001 #text change 02-Feb-2001

C:Accession: B84495

R; accession: 04932
 R; Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; N
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanKen, S.E.; Umayam, L.; Tallon, L.
 eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768. 1999

A;Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: B84495

A;Status: preliminary

A; Molecule type: DNA

A:Residues: 1-695 <STO>

A:Cross-references: GB:AE002093: NID:q4662644: PIDN:AAD26914.1: GSPDB:GN00139

C;Genetics:

A:Gene: At2g11000

A:Map position: 2

Query Match	15.2%;	Score 584;	DB 2;	Length 695;
Best Local Similarity	23.5%;	Pred. No. 1.1e-35;		
Matches 184;	Conservative 153;	Mismatches 291;	Indels 156;	Gaps 25;
QY	3	MKASVDDDSGWSLNPKEKMEKS-----NTNWVDITQDFEACRELKLGELHDKLFLGL 56		
Db	1	MQSVREDESSSPIHDSKTSISSIPSGDNNSWADVSPLLISAACSDLOEGELINDAFNL 60		
QY	57	FEAMSAIEMMDPKMDAGMIGNQNVKVLNFEQAIKOGTIKI-----KDLTLPELIGIMDT 111		
Db	61	FAAMSALEIMDPKMDSGMNST-----FYSIDEAIESGFAPVPISDSTVNVQSIIDIMDH 115		
QY	112	CFCCLLITWLEGHSLAOTVFTCLATHNPDFDTPA-MKAPALGILKICDIAFEKVNKAAVF 170		
Db	116	LLACEATWNGHSLAQVFCIIVLRPERTSSQALLHSYCRVIRATCRVVSVVSDARTN 175		
QY	171	EEDDFQSMITYGFKWANSVTDLRTVGMKVEDDMMQRRV-----KSTRSRQGEERDPEVE--- 224		
Db	176	EEDDLFTMTYGLPFSGB-EDAKGLLLNLNAVEETICRLRACKATFRRMLE--DAELEPLQ 232		
QY	225	-----LEHQOCLAVFSRVKFRVLLTVLIAATKKTSAVAEAQKLWQAADLLSAIHS--- 278		
Db	233	SNPHLEESFCKSLICIRFEKPHLHALNCMRFRPQGRGLELARKHIGYICISBELSVLSDAE 292		
QY	279	-----LHHGIIQ--AOMDITTKGBDHPITMMGFEPVWQRLPPTPRYAKIILKREEMNVFA 330		
Db	293	FRLDIFENGVNEIESSTASGRSPI--GDFPTLNKLSAPTTPRAIKLSWKAKIYYV 350		
QY	331	RLIDRIKTVCEVYNLNLNHCILDFCEFSBQSPCVLSRSLLQTTFLVDNKKVCFOTHLMQD 390		
Db	351	KLJLHNLDKICAFSLEPDLVALEFVTFQKSRPDLVARAHLQ-LILLVQDGKLYGR----- 404		

```
QY 391 MVKDARSFVDPVPLSPKCYLYNNHQAOKCIDSFVTHCVRPFCSLIQ-----IHGHNRAQ 446
Db 405 -----DFTLTICARSLADVSKNHGHTNEVILQ 433
QY 447 RDKLGHIIEFATLQDEFMTFYFNRAEKVDAALHTMLLKQBPQR-----OHLACLGTWV 500
Db 434 LNQMGIAGVG-----NMQODTSRSSKNGDKSLILNH-- 465
QY 501 LYHNLK-----IMIQYLLSGFELELYSMHEYYIYVWLSSELYAWLWSTLSRADGSQMAE 555
Db 466 IYGLEBQINWVAIRFMLGFDLDLYSPSECYMYWY--YIILWKL-----AERAR 515
QY 556 ERIW-----EQQGRSSKTKKKKVRPLSREITMSQAYQNMCGAMFKTWVAFDMGKV 610
Db 516 FRVLIVNTEERKAKRNEYSDMAREDRISLWVFLKQCCLAGLTVMIAALNEGMS 575
QY 611 RKPFEKLDSEQVREHFRAPFNSVMTPPVHYLOFKEMSDLNKYSPPQSPPELYAAASKH 670
Db 576 LKSQGPENTENEKFIQHFELLQKASLPEYDAYESFSKSTSHARLD-----YLPWVEY 627
QY 671 FQQAQKMLENIP-----NPDH--EVNRIKLVAKENFVVMKLLAGGHKKESKVPPEFDSA 723
Db 628 FHDQAKIAKOIKVGYANDPKLAETVGLKVAERNIVAVNLFC-----QDRSLKVSFBETH 693
QY 724 HKYF 727
Db 684 HPYF 687

RESULT 2
T32223
hypothetical protein T23B12.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C:Accession: T32223
R:Davidson, S.; Wohlmann, P.; Gillam, B.
submitted to the EMBL Data Library, September 1997
A:Description: The sequence of C. elegans cosmid T23B12.
A:Reference number: Z21137
A:Accession: T32223
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-814 <DAV>
A:Cross-references: EMBL:AF022982; PIDN:AAB69936.1; GSPDB:GN00023; CESP:T23B12.4
A:Experimental source: strain Bristol N2; clone T23B12
C:Genetics:
A:Gene: CESP:T23B12.4
A:Map position: 5
A:Introns: 24/1; 70/3; 468/2; 543/1; 616/1; 714/2
C:Superfamily: Caenorhabditis elegans hypothetical protein T23B12.4

Query Match 14.2%; Score 548; DB 2; Length 814;
Best Local Similarity 24.7%; Pred. No. 6.8e-33;
Matches 206; Conservative 131; Mismatches 302; Indels 196; Gaps 28;

QY 18 MPERKESNTNWVDITQDFEACRELKIGLHLLDKLFLGFEAMGAEMMDPKMDAGMIGN 77
Db 1 MEGAMESPNES-EDVSKTFKLCGLRGLGELVTEHFLSDVMSAIELGEFPMQDVG-VG- 57
QY 78 QVNRKVNFEQAOKDGIKIKDLTLPELIGIMDTFCFCLITWLGHSIAQVTFCLYIHN 137
Db 58 --SKHITLKEAISGG-LYADD--YFQLAIMDSTLAWVAWLEGSALGSTVWTVNLLSN 112
QY 138 PDFTEDPAMKAFALGILKICIDIAEKVNKAATVEE--EDFQ-SMTYGFK----- 183
Db 113 VTFVKHPVPHFASGVNLFIRNAHALNSVGNLELPEDFNPQLFHSQHWAPRRVVQL 172
QY 184 MANSVTDLRVTGMLKDVDEDDMQRVSTRSQGEERDPEVELEHQCCLAVERKVFTRVL 243
Db 173 MREQVTLIGTTG-----RKWRSAFSAQA-----YDICCavasRLEMFIML 213
QY 244 LTVL-----TAFTKE 254
```

```
Db 214 LEIIGLLVAPETEDPNFDHKVHLGDFVYHKEYEAGEAENSDEKGEESTDPTSPSEAKSE 273
QY 255 TSABAEBAQKLMVQAAD-----LLSAIHNSLHGIQAQN 287
Db 274 TTEKKE--QKSNDEATDETQNDNDENVGDEDEDDMKQFKPNFGFASILAERLCKVTKAYA 332
QY 288 DTTK-----GDHPIMMGFEPLVMQRLPPTFPYAKI1KKEEMVNYFARLIDR 335
Db 333 ETVKLGRAPDNIDGDGYLWLGAYEFTCIIRMTIPACFPKIKVPSQEAADWWVKCAER 392
QY 336 IKTVCEVNNLTNLHC--ILDFFCESEQSPCVLSRLQ--TTFLVDNKKVFGTHL----- 387
Db 393 IYHLCIVTPKTSKDLNLYLFVARTFGQACVFTSLLIQICMFPVDN-----HLCGDENR 446
QY 388 -MQDMVKDALRSFVDPVPLSPKCYLYNNHQAOKCIDSFVTHCVRPFCSLIQIHGHNRAQ 446
Db 447 SIADAVYSLSNCFPCQILDRVSPVYKQDQAQSLYVFLNHSMLAITVYSGFCNLSRQ 506
QY 447 RDKLGHIIEFATLQDEFMTFYFNRAEK--VDAALHTMLLKQBPQRQCHLACLGTWVLYHN 504
Db 507 RDRLEMALEDLGQIHS-----YAGRLBERTDEVLLSGKMTAKEQNSYSHSVAFVFNHL 561
QY 505 LRIMQYLLSGFELELYSMHEYYIYVWLSSELYAWLWSTLSRADGSQMAEERIMEEQOK 564
Db 562 LAIINHYPFELGFRMDLYVPYEFYIYWFYFSGVQAHHWRTTLERS-----QEIQLVYQ 614
QY 565 GRSSKTKTKKK-----KV-RPLSREITMSQ-----AYQNMCGAMFKTWVAFDMGKV 610
Db 615 ANPLRETNKKMLWEERCKLGEELKRRVAHAHQFSLVNLQIAISMSIDGVRLTVLIRKGI 674
QY 611 RKPFEKLDSEQVREHFRAPFNSVMTPPVHYLOFKEMSDLNK-YSPPPQSPPELYAAASK 669
Db 675 KMPKGGDAEKLRFERFEPPDSLGPPVRVDYERFKSDSGIDQMYEDKIEF--LIDQAOK 732
QY 670 HFQQAQKMLENIPNPDHEVNR-----ILKVAKNFVVMKLL 705
Db 733 SFNEAREHLEKIDN-SVEQNREMLVDVAVSFKNLLISRLQIHVAKSNIIACRVL 786

RESULT 3
S31288
MAK10 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YEL053C
C:Species: Saccharomyces cerevisiae
C>Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 21-Jul-2000
C:Accession: S31288; S30824; S50491
R:Lee, Y.J.; Wickner, R.B.
Genetics 132, 87-96, 1992
A:Title: MAK10, a glucose-repressible gene necessary for replication of a dsRNA virus of
A:Reference number: S31288; MUID:93012936; PMID:1398065
A:Accession: S31288
A:Molecule type: DNA
A:Residues: 1-733 <LEE>
A:Cross-references: EMBL:M94533; NID:g171874; PIDN:AAA34749.1; PID:g171875
R:Mulligan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Sehl, P.; Komp, C.; Wei, Y.; Taylor,
submitted to the EMBL Data Library, February 1993
A:Reference number: S30812
A:Accession: S30824
A:Molecule type: DNA
A:Residues: 1-733 <MUL>
A:Cross-references: GB:U18779; EMBL:LI0830; NID:g603625; PID:g603626
R:Dietrich, F.S.
submitted to the EMBL Data Library, December 1994
A:Description: The sequence of S. cerevisiae cosmids 8199, 8334, and 9871.
A:Reference number: S50491
A:Accession: S50491
A:Molecule type: DNA
A:Residues: 1-733 <DIE>
A:Cross-references: EMBL:U18779; NID:g603625; PID:g603626; MIPS:YEL053C
C:Genetics:
A:Gene: SGD:MAK10
A:Cross-references: SGD:S0000779; MIPS:YEL053C
```


A;Cross-references: GB:J05244
 R;Sahr, K.E.; Tobe, T.; Scarpa, A.; Laughinghouse, K.; Marchesi, S.L.; Agre, P.; Linnenbach, J. Clin. Invest. 84, 1243-1252, 1999
 A;Title: Sequence and exon-intron organization of the DNA encoding the alpha-I domain of spectrin.
 A;Reference number: A45755; MUID:90009318; PMID:2794061
 A;Accession: A45755
 A;Molecule type: DNA
 A;Residues: 1-394; 'G', 396-533 <SA2>
 A;Cross-references: GB:M29983
 R;Linnenbach, A.J.; Speicher, D.W.; Marchesi, V.T.; Forget, B.G.
 Proc. Natl. Acad. Sci. U.S.A. 83, 2397-2401, 1986
 A;Title: Cloning of a portion of the chromosomal gene for human erythrocyte alpha-spectrin.
 A;Reference number: A23533; MUID:86205962; PMID:3458204
 A;Accession: A23533
 A;Molecule type: DNA
 A;Residues: 320-450 <LIN>
 A;Cross-references: GB:M13233; NID:g182242; PIDN:AAA53103.1; PID:g182243
 R;Curtis, P.J.; Palumbo, A.; Ming, J.; Fraser, P.; Cioe, L.; Meo, P.; Shane, S.; Rovera, Gene 36, 357-362, 1985
 A;Title: Sequence comparison of human and murine erythrocyte alpha-spectrin cDNA.
 A;Reference number: A91528; MUID:86083178; PMID:3000887
 A;Accession: A91528
 A;Molecule type: mRNA
 A;Residues: 1451-1687 <CUR>
 A;Cross-references: GB:M11049; NID:9338310; PIDN:AAA60569.1; PID:g553648
 R;Speicher, D.W.; Weglarz, L.; Desilva, T.M.
 J. Biol. Chem. 267, 14775-14782, 1992
 A;Title: Properties of human red cell spectrin heterodimer (side-to-side) assembly and interaction with ankyrin.
 A;Reference number: A42872; MUID:92340516; PMID:1634521
 A;Accession: A42872
 A;Molecule type: protein
 A;Residues: 7-16; 46-55; 680-689; 1047-1056; 1921-1930 <SPI>
 R;Speicher, D.W.; Davis, G.; Marchesi, V.T.
 J. Biol. Chem. 258, 14938-14947, 1983
 A;Title: Structure of human erythrocyte spectrin. II. The sequence of the alpha-I domain.
 A;Reference number: A92408; MUID:84087888; PMID:6654896
 A;Accession: A02965
 A;Molecule type: protein
 A;Residues: 7-601 <SPC>
 R;Speicher, D.W.; Davis, G.; Yurchenco, P.D.; Marchesi, V.T.
 J. Biol. Chem. 258, 14931-14937, 1983
 A;Title: Structure of human erythrocyte spectrin. I. Isolation of the alpha-I domain and determination of its primary structure.
 A;Reference number: S13138; MUID:84087887; PMID:6654895
 A;Accession: S13138
 A;Molecule type: protein
 A;Residues: 7-92; 'X', 94-96; 110-151; 317-342; 345-366; 'Z', 368-370; 'X', 372-373; 387-434; 452-454
 R;Lusitani, D.M.; Qtaishat, N.; LaBrake, C.C.; Yu, R.N.; Davis, J.; Kelley, M.R.; Fung, J. Biol. Chem. 269, 25955-25958, 1994
 A;Title: The first human alpha-spectrin structural domain begins with serine.
 A;Reference number: A38928; MUID:95014412; PMID:7929303
 A;Contents: annotation; determination of structural domain
 C;Comment: Spectrin associates with band 4.1 and actin to form the cytoskeletal superstructure of approximately 106 residues each.
 C;Genetics:
 A;Gene: GDB:SP1A1
 A;Cross-references: GDB:119601; OMIM:182860
 A;Map position: 1q21-1q21
 C;Superfamily: spectrin alpha chain; calmodulin repeat homology; SH3 homology; spectrin/hemoglobin; actin binding; cytoskeleton; duplication; EF hand; erythrocyte; heterodimer; spectrin/dystrophin repeat homology <SP1>
 F;52-156/Domain: spectrin/dystrophin repeat homology <SP2>
 F;157-262/Domain: spectrin/dystrophin repeat homology <SP3>
 F;263-368/Domain: spectrin/dystrophin repeat homology <SP4>
 F;369-474/Domain: spectrin/dystrophin repeat homology <SP5>
 F;475-580/Domain: spectrin/dystrophin repeat homology <SP6>
 F;581-695/Domain: spectrin/dystrophin repeat homology <SP7>
 F;696-791/Domain: spectrin/dystrophin repeat homology <SP8>
 F;792-897/Domain: spectrin/dystrophin repeat homology <SP9>
 F;898-983/Domain: spectrin/dystrophin repeat homology #status atypical <SP9>
 F;984-1031/Domain: spectrin/dystrophin repeat homology <SP10>
 F;1081-1181/Domain: spectrin/dystrophin repeat homology <SP11>
 F;1182-1287/Domain: spectrin/dystrophin repeat homology <SP12>
 F;1288-1393/Domain: spectrin/dystrophin repeat homology <SP13>

F;1394-1498/Domain: spectrin/dystrophin repeat homology <SP13>
 F;1499-1604/Domain: spectrin/dystrophin repeat homology <SP14>
 F;1605-1710/Domain: spectrin/dystrophin repeat homology <SP15>
 F;1711-1816/Domain: spectrin/dystrophin repeat homology <SP16>
 F;1817-1925/Domain: spectrin/dystrophin repeat homology <SP17>
 F;1926-2032/Domain: spectrin/dystrophin repeat homology <SP18>
 F;2041-2146/Domain: spectrin/dystrophin repeat homology <SP19>
 F;2155-2257/Domain: spectrin/dystrophin repeat homology <SP20>
 F;2270-2302/Domain: calmodulin repeat homology <EF1>
 F;2313-2345/Domain: calmodulin repeat homology <EF2>

Query Match 3.5%; Score 134; DB 1; Length 2429;
 Best Local Similarity 19.4%; Pred. No. 0.37;
 Matches 175; Conservative 142; Mismatches 307; Indels 280; Gaps 48;

QY 13 GWELSMPEKMEKSNINW---VDITQD---PEEACRELKGLGELHDLKGLGFEAMSALTEM 65
 DB 1572 GNEFAMKEQLQKLEHWHLLERTNDGKLNESRQRFNTSIRDFEWFUSEAEILLAM 1631
 QY 66 MDPKMDAGMIGN-----QVNRKVINFEQAIKDTIKIKDITLPELIGIMDTFCCLITW 119
 DB 1632 KDQARDLASAGNLLKXQLLEREMLAEDALKDNLTLAEDL----- 1672
 QY 120 LEHSLAQTVTCLYIHNDFI---EDPAMKAPALGILKICIDIAREKVNKA-AVF----- 170
 DB 1673 -----LSSGTF-----NVDQIVKKDNVKNRF-LNVQELAAAHHEKLEAYALFQFFQD 1720
 QY 171 --EEDF-----QSNITYGFKMANSVTDL-----RVTGMLKQVED-----DMQRVVK 209
 DB 1721 LDDESWIEEKLIIRVSSQDYG-RDLQGVNLLKXKLEGLVAHEPAIQNVLDMAEKIK 1779
 QY 210 STRSQGEERDPEVEL-----EHQQCLAVFSRVKTRVLLTV-----LIAFTKKTSAVA 259
 DB 1780 D-KAAVGOE---ETQLRLAQFVEHWEKLEKAKARGKLEBSLEYLQPMQNAEIEEAWIN 1835
 QY 260 EAQKLMVQ--AADLLSAL-----HNSLH-----HGTAQNDYTKGDHPIMMGPEPLVNO 306
 DB 1836 EKNALAVRGDCGDTLAATQSLLMKHEALENDFAVHTRVQNVCAQG-----EDILNK 1887
 QY 307 RLDPPTPRYAKIKREEMVNYFARLDRIKTCVENVNLTLNLCILDF-FCBFSBSQSPCV 365
 DB 1888 VLHES-----QNKESIKIEALNEKTSIAKATAAWKLOLEDDYAFQEFNWKADV 1939
 QY 366 LSRSLQTTFLVDNKKVFGTHL-----MODMVKDALRSFVD---PPVLSPKCYLYN- 413
 DB 1940 EAWIADKETSUKTNGN--GADLGDFTLLAQDTLDASLQSFQBERLPEITDLKDLKLSA 1997
 QY 414 NHQAKDCIDSVTHCVVPFCSLIQIHGHNRRARQDKLGHILEEPATLO---DEPMTFYF- 469
 DB 1998 QHNSKALIEERYAALLKRWEOLEASAVHROK-----LLEKQLPQKAEIDLFEFAHK 2050
 QY 470 ---NRAEKVDAAL---HTMLIKQ--EPORQH----- 495
 DB 2051 ASALNNWCKMEENLSPFVCHVCSLNEIRLOKDHEDFLASLARAQADFKLLELDQOIK 2110
 QY 496 LG-----TWVLYHNLRIMIQYL-----LSGFEL-ELYSMHYYIY 530
 DB 2111 LGVPSSPYTLTVLELRTWKHLSDIIEERQELQKBEARQVKNFMEQEQEONASTFLQ 2170
 QY 531 WYLSEFLYAWLMTSLRADQSQAEEIMEEQGRSKTKTKKKKVPPLPREIT----- 585
 DB 2171 WILETRAYF-----LDGSLKETGTLESQLEANK-----RKOKETQAMKROLTKIVDL 2218
 QY 586 -----MSQAYQ---NMCAGMFKTVAFDMQDKVKRKPFLDSE 620
 DB 2219 GDNLEDAILLIDKYSTIGLAQWQOLYQLGLRMQHNLEQQIQAKDINGVSEETLKEFET- 2277
 QY 621 QVRYEHFAPNSVMTPPVHYLOFKE-MSDLNKYSPPPQSPPEL----- 663
 DB 2278 --LYKH-----FDENLITGLTH-KEFRSCLRLNTYLPWVERDEHEPKFEKFLDAVDGRK 2330
 QY 664 -YVAASKHFQOAKMI---LENI PNPDHEVNRILKVAKENFVVMKLLAGGHKESKVPPEF 719

Db 2331 GVSLEDY--TAFLLDKESNIKSDSEIENAFQALAEKGSYITK-----EDMKQALTPQV 2384

QY 720 DPSA 723

Db 2385 SFCA 2388

RESULT 6

S20106

hypochemical protein 2 - slime mold (Dictyostelium discoideum) retrotransposon DRE

C:Species: Dictyostelium discoideum

C:Date: 19-Mar-1997 #sequence revision 25-Apr-1997 #text_change 01-Dec-2000

C:Accession: S20106; S30209; S30211; S30207; S25646

R:Marschalek, R.; Hofmann, J.; Schumann, G.; Goesseringer, R.; Dinger, T.

Mol. Cell. Biol. 12, 229-239, 1992

A:Title: Structure of DRE, a retrotransposable element which integrates with position sp

A:Reference number: S20105; MUID:92107173; PMID:1309589

A:Accession: S20106

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1152 <MAR>

A:Cross-references: EMBL:X57034

R:Marschalek, R.; Hofmann, J.; Schumann, G.; Dinger, T.

Nucleic Acids Res. 20, 6247-6252, 1992

A:Title: Two distinct subforms of the retrotransposable DRE element in NC4 strains of D

A:Reference number: S30206; MUID:93117097; PMID:1335570

A:Accession: S30209

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: mRNA

A:Residues: 1-87,'RNR','91','S' <MA2>

A:Cross-references: EMBL:X69170; NID:g7259; PIDN:CAA48921.1; PID:g7261

A:Experimental source: retrotransposon DRE, subtype b; strain AX2

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1992

A:Accession: S30211

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: mRNA

A:Residues: 1-87,'RNR','91','S' <MAW>

A:Cross-references: EMBL:X69171; NID:g7262; PIDN:CAA48921.1; PID:g7263

A:Experimental source: retrotransposon DRE, subunit b; strain AX2

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1992

A:Accession: S30207

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-87,'XQ' <MA3>

A:Cross-references: EMBL:X68231

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1992

C:Genetics:

A:Mobile element: retroposon DRE

Query Match 3.4%; Score 132.5; DB 2; Length 1152;

Best Local Similarity 17.6%; Pred. No. 0.17;

Matches 130; Conservative 120; Mismatches 267; Indels 223; Gaps 37;

QY 18 MPKMEKSNINWVDTQDFEEACR-----ELKIGELLHDKLFLGLFAMSIAEMDPFK 69

Db 370 IPQILDPSNNCLVTRKHEDILEVARRYENLYQKRECNEDTHELLKTFNKRIEQKILD-- 427

QY 70 MDAGMIGQVNRKVLNFQAIKDGIIKIDLTLPGLIGIMDTFCFCLITWLEGH-SLAQT 128

Db 428 -----BINQPIEGYE--IRLGIEKIQEGKAPGKDGILLPTFY-----KNHINEILP 470

QY 129 VFTCLVIHNPFDIEDPAMKAFALGILKICDIAREKVNKAAVFEEDFQSMYTGFKMANSV 188

Db 471 IISKLYNH--FWNTIPKDFQGGIL--ITIIYKNKGNPNLNDYRPITLNDYKIYSKI 525

QY 189 TDLRVITGLMKVEDDMQ-----RRVKSTRSQGEERDEPV-- 223

Db 526 INNRILKULANKLIIISPQTGFVPRULLHDNIITLNSFIEIIKREINTK-----EDMEPIITF 581

QY 224 -----BLEHQOCLAVFSRVKPF-TRVLLTIVLIAPTKEKTSAAVAQKLMVQAADLLSA 274

Db 582 YDFKAFDSISHNAILRLTIAHLKPLKMLVLTIMNLINSETSVY----- 625

QY 275 IHNSLHGICQAQNDTTKGD--HPIMMGF-----EPLVNQRLLPPTPRYAKIKREE-- 324

Db 626 INNSLSKFTSKRGTKQGDPISTPIFALVTCMATIINDRCINGVTRETIKILQFADT 685

QY 325 -----MVNYFARLIDRIKTYCEV-----VNLTNLHCILDFFCFEFSEQSPCVLSRLQTT 374

Db 686 ATIAYNFMDHEL-MNEWIKKFCQATSAKINQTKSCI-----TPKWN-----TRTLY--T 732

QY 375 FLVDNKKVFGTHLMQDMVKDALRSFVDPVLPSPKCYLNNHQAQCDISFVTHCVRPFC- 433

Db 733 VIKGNERYLDFDNNGKIGKISNTISDN--TRAKLVTWNSTSS-----TYMGLLMA 782

QY 434 --SLIQIHGH--NRARQORKLGHILBEFATLODEFMTFVFNRAEKVDAALHTMLLKQE 487

Db 783 KTYALSQLTFTYINTTPOHN-----SIENNIKVFVNTKSKNSLSLQ----- 825

QY 488 PQRHACLACTGWLYVHNIRIMIQYLLSGFELELYSMHEYYIYWLSEFLYAWLMSTLSR 547

Db 826 -RRQN-----NYINGGLNLWNLKTRELAQAQWLFERYLH-----QR 860

QY 548 ADGQMAEERIMEEQKGRSSKTKKKKKVRLPSREITMSQAYQNMCMGFMKTMVAFDMD 607

Db 861 VSNTPSSYIKLWEEELKNNNNKTTTQK--NQLQHLHWQCKQAWT----- 902

QY 608 GKVRKPKFELDSEQVRYEHRFAP-----FNSVM-TppPVHYLQFKXEMSDLNKYSPPQSP 661

Db 903 -QLKTPQ-----NKQTHYEH--LPKLKKIYEDMTTQSPFH-----NKFIETPGQK 945

QY 662 ELYVAA-SKH--FOQAKMIL 678

Db 946 EIMTKINSKHLPPFKEIKKII 965

RESULT 7

E71810

type III restriction enzyme - Helicobacter pylori (strain J99)

C:Species: Helicobacter pylori

A:Variety: strain J99

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999

R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;

Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;

Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path

A:Reference number: A71800; MUID:99120557; PMID:9923682

A:Accession: E71810

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-969 <ARN>

A:Cross-references: GB:AE001563; GB:AE001439; NID:g4156027; PIDN:AA06988.1; PID:g4156025

A:Experimental source: strain J99

C:Genetics:

A:Gene: res_2

Query Match 3.3%; Score 127; DB 2; Length 969;

Best Local Similarity 17.3%; Pred. No. 0.34;

Matches 143; Conservative 125; Mismatches 271; Indels 286; Gaps 37;

QY 35 DFEACRELKGLGELHDKLFLGLFAMSIAEMDPKMDAGMIGNQV-----NRK 82

Db 9 DYQEQR-----DQILGVFK--GIDLREPENDIQRIANPVFEIGAIDKDLLENIE 56

QY 83 VLNFEQAIKDGITIKIDLTLPGLIGIMDT-----CFCLITWLEHSLAQVTCTLY---- 134

Db 57 NLRSKQITQSGVGNQSLNCDI--LMETGTGKTCFL-----ECVYALHK 100

QY 135 -IHNPDFIEDPAMKAFALGILKICDIAREKVNKAAVFEEDFQSMYTGFKMANSVTDLRV 193

Db 101 NYHLSKFIVLAPNSAIKGLVLSIETRE-----FEKSEYSNHL-- 140

QY 194 TGLMKDVEDDMQRRYKSTRSQGEERDPEVELEHQOCLAVFSRVKPTRVLLTV----- 246

Db 141 -----ESYEDVERFILAS-----NHKCCVLVMTFSAFNKEKNTINQSCLEN 181

R.Wu, X.

submitted to the EMBL Data Library, November 1995

A;Description: The sequence of C. elegans cosmid F20D12.

A;Accession: Z18462

A;Accession: T16111

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-778 <WUX>

A;Cross-references: EMBL:U40933; NID:gl072142; PID:gl072145; PIDN:AAA81674.1; CESP:F20D12

C;Genetics:

A;Gene: CESP:F20D12.4

A;Introns: 98/3; 284/2; 465/2; 735/3

C;Superfamily: Caenorhabditis elegans hypothetical protein F20D12.4

Query Match 3.2%; Score 123.5; DB 2; Length 778;
Best Local Similarity 18.9%; Pred. No. 0.45;
Matches 115; Conservative 89; Mismatches 188; Indels 215; Gaps 28;

QY 13 GWELSMPEK---MEKSNTNW---VDITQDFEACRELKL-----GEL 48
DB 144 GFDMSVPLKFAAEVSLVSYNCRQLSADYRAMNVPKLSKQKCGDRTNVSFVNVGSV 203
QY 49 LHDKLFG-LFPAIAIEMMDPKMDAGMTGNQNRKVLN-PEQAI---KQGT-IKIKDLTL 102
DB 204 EDQKMLNETLSAMNIGQLPERLDAWKI-----VILNVFCEAIVASRDGVDVIVDNP 257
QY 103 PELIGIMTCFCCLITWEGHSLAQTVETCLYIHPDPFIEDPKAPALGILKICIDIARE 162
DB 258 PD-----QTRF---LINQKPRGKKD-----KTIDVAKV 282
QY 163 KVNKAADF-----EEDFQSMTVGFKMANSVTLRLVGMKDV-----E 201
DB 283 LESMEVFTKLHSLVHSHLLDQTKTFTSMIGSVIEQLITMILKDVIAAAPTETAD 342
QY 202 DMQRRVSTRSQGEERDPEVELEHQCLAVFS-RVRFTRVLLTVLIAFTKKETSAVAE 260
DB 343 EDQEMFINLLQI-----GEVFVERMKELGPFQSKAKLLFTLDTDTFTVTRRCFAIVSK 395
QY 261 AOKLM-----VQAADILSAHNSLH-----GIAQNDTTKGDH 294
DB 396 ANKLINETYDKLVTVGVDSADIKDILLAKAHATHAEHFAKEYGKDLGRLSHNSDSQF-- 453
QY 295 PTMGFEPLVNRQLPPTFPYAKIKKEEMVNYEARLIDRKT--VCE----- 341
DB 454 -----PFFFAFOKTVSESTINFNLRLDNVKAFAACEDEGARAKIALT 497
QY 342 VVNLNLHCLD-----FFCFSEQSPCVLSRLSLQTTFLVDNKKVFG 384
DB 498 AENIVRLVILTPRKHAELFSSIPNMAAIFYNNCHYISHCMTMGFEASG---DNQKTL 554
QY 385 THLMQDMVKDALRSFVDPVLPSPKCYLYNNHQAOKCIDSFVTHCVRPFCSLIQIHGNRA 444
DB 555 EPLLDSDV-IRLRIV-----AADCKEKLTRCREMTAYLEDH----- 591
QY 445 RQRDKLGHILEEFATLQDEFMTFFYNRAEKVDAAALHTMLLKQEPQORQHLACIGTWLVY-- 502
DB 592 ---SIFELPASYKTKNTFAA-----AEEMSESADILVPREP--KLIKCLAACLLH 641
QY 503 ---HNLR 506
DB 642 LIAKNLR 648

RESULT 10

T21982

hypothetical protein F39B1.1 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000

C;Accession: T21982; T22312

R;Lightning, J.

submitted to the EMBL Data Library, February 1996

A;Reference number: Z19497

A;Accession: T21982

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-1607 <WIL>

A;Cross-references: EMBL:Z69660; PIDN:CAA93489.1; GSPDB:GN00028; CESP:F39B1.1

A;Experimental source: Clone F39B1

R;Thomas, K.

submitted to the EMBL Data Library, March 1996

A;Reference number: Z19545

A;Accession: T22312

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-1607 <WIL>

A;Cross-references: EMBL:Z69903; PIDN:CAA93776.1; GSPDB:GN00028; CESP:F39B1.1

A;Experimental source: clone F46F2

C;Genetics:

A;Gene: CESP:F39B1.1

A;Map position: X

A;Introns: 15/2; 65/2; 134/1; 190/1; 261/3; 374/2; 417/3; 472/2; 512/1; 661/3; 715

C;Superfamily: Hsc2 phosphatidylinositol 3-kinase; protein kinase C C2 region homology

Query Match 3.2%; Score 122.5; DB 2; Length 1607;
Best Local Similarity 19.4%; Pred. No. 1.5;
Matches 105; Conservative 83; Mismatches 177; Indels 175; Gaps 24;

QY 154 LKICDIAREKVNKAAVFEEDFQSMYTG-----FKMANSVTLRLVTG 195
DB 465 LESNLMKSVKYSIVVDKEDVEN-TLGLHASEMSQVEIAFNDGSTLKLSSSQRVKQVI 523
QY 196 ML-----KDVDDMQRRVKSTRSQGEERDPEVELEHQCLAVFSRVKTRVLLT 245
DB 524 MLLCKLGHIVPEKLYNEMQKYLASTTQDLVHRNDFLRHISFLEYLCRCYCRVSRYNIP 583
QY 246 VLIAPTKETSAVAFAOKLMVQAADLLSAHN-----SLHGIQAQN-- 287
DB 584 FLQIITKEKVLKMDFLQI---MLNSVHSIPEHWOSQYSEFYNSLDLXHGTVLDGF 639
QY 288 -----DTTKGDHPTMMGPEPLVNRQLLPPTPR-----YAKIKKEEMVNY-----FAR 331
DB 640 SNKPKTITKNDH-----FFRIPLDIYAK-FKRLNLCOYPRETRIVVS 681
QY 332 LIDRIKTVCEVVNLTNLHCILDFFCE-----FSEQSPCVLSRLSLQTTFLVDNKKV 383
DB 682 ISGTVRNSQAANEYNDIVMLGYSVPLYDENLFMRQGLPLFLTLILKKQPM-----KPF 738
QY 384 GTHLMQDMVKA-----LRSFVDPVLPSPKCYLYNNHQAOKCI-DSFVTHCVRP---FCS 434
DB 739 GPY---PYIKDARDILIMSP---KIMDTETFPNVVIDMQCIPQDFAPLDIETQEYLLE 792
QY 435 LIQ-----IHGHNH-----ARQD-----KLGHILEEFATL 460
DB 793 LIENQDTSTLEDDODLLWQKELHLTNQPEALPLVLSLQDWSFGFVMEVYQLLEWAPL 852
QY 461 QDEFMTFFYNRAEKVDAALHTMLLKQEPQOR-----OHLACIGTWLVYNNHILMTOYL-- 512
DB 853 RPEI-----AMEFLPKYPDERIRAHAVQSLARGSTDFLYHTTIPQFTEALRF 899
QY 513 -----LSGFELEH---YSMEHYVYIYWLSEFLYAWLMSTLSRADGSQMAERIMEEQ 563
DB 900 ELYEKSALADFILELSFVSLDFTFEIYWLQQRVDHCAVDDLPYAIRCNQLQQRKIDHEH 959

RESULT 11

T30336

nuclear/mitotic apparatus protein - African clawed frog

C;Species: Xenopus laevis (African clawed frog)

C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000

C;Accession: T30336

R;Merdes, A.; Ramvar, K.; Vechio, J.D.; Cleveland, D.W.

Cell 87, 447-458, 1996

A;Title: A complex of NuMA and cytoplasmic dynein is essential for mitotic spindle assem

A;Reference number: Z20828; MUID:97053784; PMID:8898198

A;Accession: T30336

A;Status: preliminary; translated from GB/EMBL/DBDJ

Db 1082 LEAPQWYNTLRTVCDKWIRMKCKVEMDARRSPMEHE-----NKEWTLRTTELLAARPQMP 1137

QY 638 PPVHYLQFKMSDLN-----KYSPPPQPELYVVAASKHFOQAQKMLLENIPN-----PDHE 687

Db 1138 SSLRLQRPQREVLQAFNAKWKSGSKSEITVNAQLEQSTQKKFIRNIERYKGIPIYE 1197

QY 688 VNRLLKVAKPNFVVMKLLAGGKKESKVPPEFDFSAHKYFVPVK 731

Db 1198 L--LIKAARLQVNV-----VSEYKLTPOSEYTFSRIFDLIK 1232

RESULT 13

TI13057

KIAA0729 protein - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000

C:Accession: T13057; T08741

R:Nagase, T.; Ishikawa, K.; Suyama, M.; Kikuno, R.; Miyajima, N.; Tanaka, A.; Kotani, H.

DNA Res. 5, 277-286, 1998

A:Title: Prediction of the coding sequences of unidentified human genes. XI. The complete

A:Reference number: Z17595; MUID:99087487; PMID:9872452

A:Accession: T13057

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1196 <NAG>

A:Cross-references: EMBL:AB018272; NID:G3882178; PIDN:BAA34449.1; PID:G3882179

A:Experimental source: Brain

R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, March 1999

A:Reference number: Z16471

A:Accession: T08741

A:Molecule type: mRNA

A:Residues: 279-1196 <NAG>

A:Cross-references: EMBL:AL050092

A:Experimental source: adult uterus; clone DKFZp586G0518

C:Genetics:

A:Note: KIAA0729; DKFZp586G0518.1

Query Match 3.1%; Score 119; DB 2; Length 1196;

Best Local Similarity 19.3%; Pred. No. 1.8;

Matches 124; Conservative 99; Mismatches 235; Indels 186; Gaps 31;

QY 97 IKDLTLPPLIGIMDTFCCLITWLBGHSLAQVFTCLYHNPDFIEDPAMKAFALGILKI 156

Db 563 MREELEIDIKFKKTTISCYRLCDGRSCWTLIS-----AFRILE 604

QY 157 CDIAREKV--NKAAYFEEDFSQMTYGFPMANSVTDLRVGMLKQVEDDMQRRVKSTRSR 214

Db 605 SDEDRLLVFNRLTMTESFNTL--HMVYHEATACTVGDVLVLLSIFLSVLKSTRPY 661

QY 215 QGEERDPEVELEHQOCLAVFSVKETRVLLTVLIAFTKKE--TSVAEAKQ--LMVQAADL 271

Db 662 LQRKDVQKALIQWQE-----RIEFAHKLLTLNLSYSPPELNACIDVLKELVLLSPHDF 715

QY 272 LSAI-----HNSLHHGTAQNDTTKGDPHMMG--FEPLVNORLL-----PP----- 311

Db 716 LHTLVFPLQHNHCYHH-----SNIPMSLGYPFCRENKILGKSNIRPPEL 765

QY 312 ---TPFYAKIIKREEMVNYFARLIDRIKTVCEVNVNLNHLCHILDFPCE--FSEQSPCVLS 367

Db 766 NMCLLPTTWETSKGDDV-YDRMLLDYFYSYHQFIHL--CRVAINCEKETE----- 814

QY 368 RSLQTTFLVDNKKV-----FGTHLMQD---MVKDARSFVDPPVLSL--KC----- 409

Db 815 -TLVLSLVLAYEGLPLHALFPKLMTELCTQTSAMSKNCIKLCEBDFVFAEYIKILMD 873

QY 410 ---LYNNHQAQCDIDSFVTHCVRPFCSLIQIHGHRARQDKLG---HILEEPAFLQD 462

Db 874 ERTFLNNN-----IVTFMTH---FLKVKQSVSEANCANLSTLITNLISQVQNLOS 924

QY 463 EFTMYFYNRAE--KVDAALH-----TMLKQEPQRQHLACLTGTVWLYHNLRLIMIQYLLS 514

Db 925 DFS---NRVEISKASASINGDLRALALLSVHTPKQ----- 957

QY 515 GFELELYSMHEYYIYWVLSFLYANLWMTLSRADGSQMAEERIMEEQOKGRSSKTKKK 574

Db 958 -----LNPAIPTTQELLSKTRCTLQQNSLQEQAKER---KTDD 996

QY 575 KKVRLGREITMSQAYQNM--CAGMFKTWVAFDMGDKVRKPKFELDSQVRYEHRFAPFN 632

Db 997 EGATPIKERRRVSSDEEHTVDSICSDMTETR-----EVLTFSTSDNE-----TRDS 1043

QY 633 SVMTPTPPVHYLQFKMSDLNKYS---PPQSPFELYVAAASKHFOQ 673

Db 1044 SIIDPGTEQDLSPENSSVKEYRMEVPSFSEDMSNIRSOAAEE 1087

RESULT 14

T02576

hypothetical protein At2g39260 [imported] - Arabidopsis thaliana

N:Alternate names: hypothetical protein T16B24.10

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-feb-2001

C:Accession: T02576; B84815

R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.

submitted to the EMBL Data Library, August 1998

A:Description: Arabidopsis thaliana chromosome II BAC T16B24 genomic sequence.

A:Reference number: Z14679

A:Accession: T02576

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1029 <ROU>

A:Cross-references: EMBL:AC004697; NID:G3402671; PID:G3402679

A:Experimental source: cultivar Columbia

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; I.

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.C.

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.W.; Venter, J.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: B84815

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1029 <STO>

A:Cross-references: GB:AB002093; NID:G3402679; PIDN:AAC28982.1; GSFDB:GN00139

C:Genetics:

A:Gene: At2g39260; T16B24.10

A:Map position: 2; 104/3; 147/2; 205/2; 232/3; 330/3; 396/3; 430/3; 514/3; 587/3; 644/3; 70

A:introns: 72/3; 104/3; 147/2; 205/2; 232/3; 330/3; 396/3; 430/3; 514/3; 587/3; 644/3; 70

Query Match 3.1%; Score 118.5; DB 2; Length 1029;

Best Local Similarity 18.8%; Pred. No. 1.6;

Matches 87; Conservative 82; Mismatches 132; Indels 163; Gaps 23;

QY 21 KMEKSNNTWV---DITQDFEACRELKLGELLHDKLGLFEAMS-AEMMDPKMDAGMI 75

Db 108 QMEKENAKLVNAKGLSEDSASSYKLRKS---YDHLRYNTSSLAELDMQPPVMP--- 160

QY 76 GQVNRKVLNFEQATKGTIKIKDLTLPGLIGIMTFCCLITWLEHSLAQTVTCLVI 135

Db 161 -EDGTRITAGDEASPSGT--VKUTSVPEPI-----WDEDT--KTFECL-- 201

QY 136 HNPDIEDPAMKAFALGIL-----KICDIAREKVNKAA---VFBEEDFSQMTYGFK 183

Db 202 -----PDLRAFVPAVLGEAEPKSNEQSAKAKELSESSESVENQQTEDTTEVSA 253

QY 184 MANSVTDLRVGMLKQVEDDMQRRVKSTRSQGEERDPEVELEHQ---OCLAVFSRVKF 239

Db 254 DSASMDRNSNAEQPKKEEVEKEKAKDTKKGKGEKDEKKMEHEKKGKSLDV---ANF 310

QY 240 TFLV-----LTVLIAFTKKTETSAVAEAKLWQAADLLSAIHNSLHHGICA 285

Db 311 EKLRLQPGCVSRDLIDQLTVEYCYLNSKTN-----RKLKLVKA--LFFNVRTSL 357

QY 286 QNDTTKGDPHMMGPEPLVNQRLPPTFPYAKIIKREEMVNYFARLIDRIKTVCEVNVL 345

Db 358 -----ELLAYGRMV-----A 368
 QY 346 TNLHCLIDFFCFEFSQSPVLSRLQTFE-LVDNK-----KVFGTHLMQDMVKDAL 396
 Db 369 TLASCKDI-----PSMLVQ-MLEDEFNSLVHKKQOMIETKIRNIRFIGELCKPK- 418
 QY 397 RSFVDPVPLPKCYLYNNHQAKDCIDSVFTHCVRFCSLIQIHG 440
 Db 419 ---IVPAGLVFSC-----LKACLDFTFHNDIVACNLLETCG 452

RESULT 15
 S02041
 N/Dystrophin, muscle - chicken
 C/Species: Gallus gallus (chicken)
 C/Date: 07-Sep-1990 #sequence revision 27-Jun-1994 #text_change 16-Jul-1999
 C/Accession: S02041; S02013; S71487
 R/Lemaire, C.; Heilig, R.; Mandel, J.L.
 Nucleic Acids Res. 16, 11815-11816, 1988
 A/Title: Nucleotide sequence of chicken dystrophin cDNA.
 A/Reference number: S02041; MUID:89098331; PMID:3062582
 A/Accession: S02041
 A/Status: translation not shown
 A/Molecule type: mRNA
 A/Residues: 1-3660 <LEM>
 A/Cross-references: EMBL:X13369; NID:G63369; PIDN:CAA31746.1; PID:G63370
 A/Note: 1869-His, 1885-Arg, and sequences lacking 1171-Met were also found
 R/Lemaire, C.; Heilig, R.; Mandel, J.L.
 EMBO J. 7, 4157-4162, 1988
 A/Title: The chicken dystrophin cDNA: striking conservation of the C-terminal coding and
 A/Reference number: S02013; MUID:89210800; PMID:3072195
 A/Accession: S02013
 A/Status: nucleic acid sequence not shown
 A/Molecule type: mRNA
 A/Residues: 1-3573, 'HA', 3576-3660 <LEM2>
 R/Heilig, R.; Lemaire, C.; Mandel, J.L.
 Nucleic Acids Res. 15, 9129-9142, 1987
 A/Title: A 230kb cosmid walk in the Duchenne muscular dystrophy gene: detection of a con
 A/Reference number: S09071; MUID:88067745; PMID:2825128
 A/Accession: S71487
 A/Molecule type: DNA
 A/Residues: 222-281 <HEI>
 C/Comment: Dystrophin is proposed to play a role in anchoring the cytoskeleton to the pl
 C/Comment: Defects in dystrophin are responsible for the Duchenne/Becker muscular dystro
 C/Superfamily: dystrophin; alpha-actinin actin-binding domain homology; spectrin/dystro
 C/Keywords: actin binding; calmodulin binding; cytoskeleton; leucine zipper; membrane-as
 F:1267/Domain: alpha-actinin actin-binding domain homology <ACT>
 F:253-327/Region: hinge
 F:340-449/Domain: spectrin/dystrophin repeat homology <SP01>
 F:450-558/Domain: spectrin/dystrophin repeat homology <SP02>
 F:560-669/Domain: spectrin/dystrophin repeat homology <SP03>
 F:670-719/Region: hinge
 F:720-830/Domain: spectrin/dystrophin repeat homology <SP04>
 F:838-936/Domain: spectrin/dystrophin repeat homology <SP05>
 F:940-1047/Domain: spectrin/dystrophin repeat homology <SP06>
 F:1049-1156/Domain: spectrin/dystrophin repeat homology <SP07>
 F:1158-1265/Domain: spectrin/dystrophin repeat homology <SP08>
 F:1267-1369/Domain: spectrin/dystrophin repeat homology <SP09>
 F:1374-1479/Domain: spectrin/dystrophin repeat homology <SP10>
 F:1480-1570/Domain: spectrin/dystrophin repeat homology #status atypical <SP11>
 F:1572-1678/Domain: spectrin/dystrophin repeat homology <SP12>
 F:1680-1784/Domain: spectrin/dystrophin repeat homology <SP13>
 F:1787-1877/Domain: spectrin/dystrophin repeat homology #status atypical <SP14>
 F:1878-1984/Domain: spectrin/dystrophin repeat homology <SP15>
 F:1986-2103/Domain: spectrin/dystrophin repeat homology <SP16>
 F:2105-2211/Domain: spectrin/dystrophin repeat homology <SP17>
 F:2213-2319/Domain: spectrin/dystrophin repeat homology <SP18>
 F:2323-2419/Domain: spectrin/dystrophin repeat homology <SP19>
 F:2420-2467/Region: hinge
 F:2468-2574/Domain: spectrin/dystrophin repeat homology <SP20>
 F:2576-2683/Domain: spectrin/dystrophin repeat homology <SP21>
 F:2685-2799/Domain: spectrin/dystrophin repeat homology <SP22>

F:2801-2928/Domain: spectrin/dystrophin repeat homology <SP23>
 F:2930-3037/Domain: spectrin/dystrophin repeat homology <SP24>
 F:3038-3075/Region: hinge
 F:3082-3089/Domain: WW repeat homology <WW1>
 F:3079-3357/Region: cysteine-rich
 F:3481-3502/Region: leucine zipper motif
 F:3547-3568/Region: leucine zipper motif

Query Match 3.1%; Score 117.5; DB 1; Length 3660;
 Best Local Similarity 17.7%; Pred.No.11;
 Matches 132; Conservative 113; Mismatches 229; Indels 273; Gaps 35;

QY 13 GWELSMPEKMEKSNINWVDIT----- 33
 Db 1646 GKESLVEDKLSLLNSNWTAVTSRAEWNLLMEYQKHEAFDQKVANTTWIYRAEILLD 1705
 QY 34 -----ODFEACRELKLGEL--LHDKLFGLFEMSAIEMM-----DPKMDAG 73
 Db 1706 ESDKQKPOQKEETLKRLLK-AELNDMPKVDVSRD--QAVDLMTNRGDHCRKVIEPKL--- 1759
 QY 74 MIGNQVNRKVLNFEQAIKDGTIKIKDLTLPELIGIMDTFCCLITWLGHSLAQVFTCL 133
 Db 1760 ---SELNHRFAAISQRIKSG----- 1776
 QY 134 YIHNPDFIEDPAMKAFALGILKIIDIAREKVNKAAVFEEDF-----QSMTY 180
 Db 1777 ---KFFIPLAKELQGFDFDIQKLEPLEVEIQQGNLKEEDFNKMDSEDDSTVKELLQR 1832
 QY 181 GFKMANSVTDLRVTCMLKDVDDMQRRVKSTRSQGGEERDEVELEHQOCLAVSRVKFT 240
 Db 1833 GDTLQKRTIDRKRBEIKIKOQLLTQTHNALKDLRSORRKALETISHW---YQYKRQA 1888
 QY 241 RVLLTLVLIATFKETSAV--AEAOKIMVQAADL-----LSAIHNSLHHGQAQNDTKG 292
 Db 1889 DDLMTWLDIEKKLASLPDHKDEQKLKEIGELKKEKEDLNNAV---QAEKLSKDG 1942
 QY 293 DHPIMMGPELVNQRLLPPTFPFYAKIIRKREVMNYFARLIDRINKTVCVVNLTNLHCIL 352
 Db 1943 -----AAKAVEPTLVQLSK--RWRDFESKPAQF-----RLNYAQIQITVL 1980
 QY 353 DFCFESQSPCVLSRSL-LQTTFLVDNKKVFGTHLMQDM--VKDALRSFVDPVPLSPKC 409
 Db 1981 -----EDTTFVMTESMTVETVPSTYLAELIQLQALSEVERLNS---PVL----- 2025
 QY 410 YLYNNHOAKDCIDSF-VTHCVRPF--CSLIGIHGHNRAORDKLGHIL---EEFATLQD 462
 Db 2026 -----QAKCEDLLKQBECLKNKDC-LGRLOQHIDIHKKTPALQSATPRETANIQD 2078
 QY 463 EFMFTYFNRAEKVDAALHTMLLKQEPQPOHIACLGTWLVYHNLRIMIQYLLSGFELELYS 522
 Db 2079 K-LTQLNSQWBEKVNMYRDRQARFDSKE-----KWLRFH-----CEMK 2117
 QY 523 MHEYYYIYWLSEFLYAWLMST---LSRA-----DGSQMAEERIMWEOOKGRSSKTKKK 574
 Db 2118 FNE-----WLTEBEKLSRAQIEAGDVGHVTKQFLQELQDQIGRQQTIVK 2163
 QY 575 KKVRLPLGREITMSQAYQNMCAGMFKTMVAFMDGKVRKPKF-----ELDSEOVRYEH 626
 Db 2164 -----TLNVTGEIILIQSSAA-----DANVLKEQLNLTNRWQETCRQLVEKRK 2207
 QY 627 RFAPNSVMTPPPVHYLQFKEMSDLNK 653
 Db 2208 RIEBEKNILS-----EFQE--DLNK 2225

Search completed: August 10, 2004, 21:52:01
 Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 21:47:26 ; Search time 13 Seconds
(without alignments)
2935.956 Million cell updates/sec

Title: US-10-001-857-145

Perfect score: 3849

Sequence: 1 MVMKASVDDDDSGWELSMPE.....KVPPEFDSAHKYPVVKLV 733

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	270.5	7.0	733	1	MK10 YEAST
2	134	3.5	2418	1	SPCA HUMAN
3	127	3.3	733	1	SAB2 HUMAN
4	127	3.3	733	1	SAB2 MOUSE
5	127	3.3	1089	1	NMD2 YEAST
6	125.5	3.3	1603	1	VIT3 CAEL
7	125.5	3.3	1603	1	VIT4 CAEL
8	121.5	3.2	962	1	IF3A MAIZE
9	121.5	3.2	1603	1	VIT5 CAEL
10	117.5	3.1	3660	1	DMD CHICK
11	116	3.0	736	1	BAC1 HUMAN
12	115	3.0	1187	1	ALIA MOUSE
13	114	3.0	1679	1	YIO9 YEAST
14	113.5	2.9	1957	1	SPOF SCHPO
15	113	2.9	2787	1	TEL1 YEAST
16	113	2.9	3744	1	TRAL YEAST
17	111.5	2.9	1332	1	IKAP HUMAN
18	111	2.9	607	1	DOR1 YEAST
19	111	2.9	1067	1	EG52 XENLA
20	110.5	2.9	2415	1	SPCA MOUSE
21	109.5	2.8	564	1	M12 STRPY
22	109.5	2.8	1066	1	KL61 DROME
23	108.5	2.8	717	1	HS90 BRUPA
24	108.5	2.8	718	1	RH88 HUMAN
25	108.5	2.8	1442	1	DPO3 UREPA
26	108.5	2.8	2672	1	GCN1 YEAST
27	108	2.8	837	1	RA50 METTH
28	108	2.8	1953	1	BN11 YEAST
29	107.5	2.8	886	1	RA50 SULAC
30	107	2.8	1065	1	SE88 YEAST
31	106	2.8	900	1	IF38 ARATH
32	106	2.8	925	1	Y562 HUMAN
33	106	2.8	1015	1	CND3 HUMAN

34 105.5 2.7 766 1 AMD3 MOUSE
35 105.5 2.7 1507 1 SIMA_DROME
36 105 2.7 732 1 HS9A_MOUSE
37 105 2.7 1312 1 RA50 YEAST
38 105 2.7 8797 1 SNE1 HUMAN
39 104.5 2.7 725 1 YKZ8 YEAST
40 104.5 2.7 767 1 AMD3_HUMAN
41 104.5 2.7 858 1 YNUL YEAST
42 104.5 2.7 1125 1 YB62_SCHPO
43 104.5 2.7 1134 1 ALIA HUMAN
44 104.5 2.7 1385 1 RRPO_PIAMV
45 104.5 2.7 4644 1 DYHC_MOUSE

O08739 mus musculus
Q24167 drosophila
P07901 mus musculus
P12753 saccharomyc
Q8nf91 homo sapien
P36114 saccharomyc
Q01432 homo sapien
P40164 saccharomyc
O14248 schizosacch
P98196 homo sapien
Q07518 plantago as
Q9jhu4 mus musculus

ALIGNMENTS

RESULT 1
MK10 YEAST STANDARD; PRT; 733 AA.
AC Q02197;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Glucose repressible protein MAK10.
GN MAK10 OR YEL053C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93012936; PubMed=1398065;
RT Lee Y.-J., Wickner R.B.;
RT "MAK10, a glucose-repressible gene necessary for replication of a
RT dsRNA virus of Saccharomyces cerevisiae, has T cell receptor
RT alpha-subunit motifs.";
RL Genetics 132:87-96(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288c / AB972;
RC MEDLINE=97313264; PubMed=9169868;
RA Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,
RA Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,
RA Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G., Lew H.,
RA Hunnicke-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,
RA Lin D., Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P.,
RA Oh C., Petel P.X., Roberts D., Sehl P., Schramm S., Shogren T.,
RA Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome V.";
RL Nature 387:78-81(1997).
CC -!- FUNCTION: Has a role in the propagation of L-A and M viruses,
CC perhaps in the viral assembly. It is apparently directly needed
CC for optimum respiration.
CC -!- INDUCTION: Glucose-repressed.
CC -!- SIMILARITY: SOME, TO T-CELL RECEPTOR ALPHA SUBUNIT VARIABLE
CC REGIONS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M94533; AAA34749.1; -;
CC EMBL; U18779; AAB64989.1; -;
CC FIC; S31288; S31288.
CC GerMOnline; 139057; -;
CC SGD; S0000779; MAK10.
CC GO; GO:0004042; P:amino-acid N-acetyltransferase activity; IDA.
CC GO; GO:0006474; P:N-terminal protein amino acid acetylation; IDA.

RN VARIANT EL2 SER-28, AND VARIANT HPP ARG-48.
 RX MEDLINE=91346849; PubMed=1878597;
 RA Floyd P.B., Gallagher P.G., Valentino L.A., Davis M., Marchesi S.L.,
 RA Forget B.G.;
 RT "Heterogeneity of the molecular basis of hereditary
 RT pyropoikilocytosis and hereditary elliptocytosis associated with
 RT increased levels of the spectrin alpha I/74-kilodalton tryptic
 RT peptide.";
 RL Blood 78:1364-1372(1991).
 [12]
 RN VARIANT EL2 SER-45.
 RX MEDLINE=89323468; PubMed=2568862;
 RA Lecomte M.-C., Garbarz M., Grandchamp B., Feo C., Gautero H.,
 RA Devaux I., Bournier O., Gaiand C., D'Auriol L., Galibert F.,
 RA Sahr K.E., Forget B.G., Boivin P., Dherym D.;
 RT "Sp alpha I/78: a mutation of the alpha I spectrin domain in a white
 RT kindred with HE and HPP phenotypes.";
 RL Blood 74:1126-1133(1989).
 [13]
 RN VARIANT EL2/HPP PRO-207.
 RX MEDLINE=92176375; PubMed=1541680;
 RA Gallagher P.G., Tse W.T., Coetzer T., Lecomte M.-C., Garbarz M.,
 RA Zarkowsky H.S., Baruchel A., Ballas S.K., Dherym D., Palek J.,
 RA Forget B.G.;
 RT "A common type of the spectrin alpha I 46-50a-kD peptide abnormality
 RT in hereditary elliptocytosis and pyropoikilocytosis is associated
 RT with a mutation distant from the proteolytic cleavage site. Evidence
 RT for the functional importance of the triple helical model of
 RT spectrin.";
 RL J. Clin. Invest. 89:892-898(1992).
 [14]
 RN VARIANT VAL-1857.
 RX MEDLINE=9325053; PubMed=8486776;
 RA Wilmotte R., Marechal J., Morle L., Baklouti F., Philippe N.,
 RA Kastally R., Kotula L., Delaunay J., Alloisio N.;
 RT "Low expression allele alpha 1B1 of red cell spectrin is associated
 RT with mutations in exon 40 (alpha V/41 polymorphism) and intron 45 and
 RT with partial skipping of exon 46.";
 RL J. Clin. Invest. 91:2091-2096(1993).
 [15]
 RN VARIANT EL2 BARCELONA PRO-469.
 RX MEDLINE=93372367; PubMed=8364215;
 RA dalla Venezia N., Alloisio N., Forissier A., Denoroy L., Aymersch M.,
 RA Vives-Corrons J.L., Besludch J., Besson I., Delaunay J.;
 RT "Elliptocytosis associated with the alpha 469 His-->Pro
 RT mutation in spectrin Barcelona (alpha I/50-46b).";
 RL Blood 82:1661-1665(1993).
 [16]
 RN VARIANT CAGLIARI GLY-2024.
 RX MEDLINE=94043025; PubMed=8226774;
 RA Sahr K.E., Coetzer T.L., Moy L.S., Derick L.H., Chishiti A.H.,
 RA Jarolim P., Lorenzo F., Miraglia del Giudice E., Iolascon A.,
 RA Gallanello R.;
 RT "Spectrin Cagliari: an Ala-->Gly substitution in helix 1 of beta
 RT spectrin repeat 17 that severely disrupts the structure and self-
 RT association of the erythrocyte spectrin heterodimer.";
 RL J. Biol. Chem. 268:22656-22662(1993).
 [17]
 RN VARIANT EL2 CULOZ VAL-46, AND VARIANT EL2 LYON PHE-49.
 RX MEDLINE=90347052; PubMed=2384601;
 RA Morle L., Roux A.-F., Alloisio N., Pothier B., Starck J., Denoroy J.,
 RA Morle F., Rudigoz R.-C., Forget B.G., Delaunay J., Godet J.;
 RT "Two elliptocytogenic alpha I/74 variants of the spectrin alpha I
 RT domain. Spectrin Culoz (GGT-->GTT; alpha I 40 Gly-->Val) and
 RT spectrin Lyon (CTR-->TRT; alpha I 43 Leu-->Phe).";
 RL J. Clin. Invest. 86:548-554(1990).
 [18]
 RN VARIANT EL2 JENDOUBA GLU-791.
 RX MEDLINE=92345619; PubMed=1638030;
 RA Alloisio N., Wilmotte R., Morle L., Baklouti F., Marechal J.,
 RA Ducluzeau M.-T., Denoroy L., Feo C., Forget B.G., Kastally R.,
 RA Delaunay J.;

RT "Spectrin Jendouba: an alpha II/31 spectrin variant that is
 RT associated with elliptocytosis and carries a mutation distant from
 RT the dimer self-association site.";
 RL Blood 80:809-815(1992).
 [19]
 RN VARIANT EL2 TUNIS TRP-41.
 RX MEDLINE=89323436; PubMed=2568861;
 RA Morle L., Morle F., Roux A.F., Godet J., Forget B.G., Denoroy L.,
 RA Garbarz M., Dherym D., Kastally R., Delaunay J.;
 RT "Spectrin Tunis (Sp alpha I/78), an elliptocytogenic variant, is due
 RT to the CGG-->TGG codon change (Arg-->Trp) at position 35 of the
 RT alpha I domain.";
 RL Blood 74:828-832(1989).
 [20]
 RN VARIANT EL2 GENOVA TRP-34.
 RX MEDLINE=94250920; PubMed=8193371;
 RA Perrotta S., del Giudice E.M., Alloisio N., Sciaratta G., Pinto L.,
 RA Delaunay J., Cuttillo S., Iolascon A.;
 RT "Mild elliptocytosis associated with the alpha 34 Arg-->Trp mutation
 RT in spectrin Genova (alpha I/74).";
 RL Blood 83:3346-3349(1994).
 [21]
 RN VARIANT EL2 ANASTASIA THR-45.
 RX MEDLINE=95290423; PubMed=7772539;
 RA Perrotta S., Iolascon A., de Angelis F., Pagano L., Colonna G.,
 RA Cuttillo S., del Giudice E.M.;
 RT "Spectrin Anastasia (alpha I/78): a new spectrin variant (alpha 45
 RT Arg-->Thr) with moderate elliptocytogenic potential.";
 RL Br. J. Haematol. 89:933-936(1995).
 CC -!- FUNCTION: Spectrin is the major constituent of the cytoskeletal
 CC network underlying the erythrocyte plasma membrane. It associates
 CC with band 4.1 and actin to form the cytoskeletal superstructure of
 CC the erythrocyte plasma membrane.
 CC -!- SUBUNIT: Composed of nonhomologous chains, alpha and beta, which
 CC aggregate side-to-side in an antiparallel fashion to form dimers,
 CC tetramers, and higher polymers.
 CC -!- DISEASE: Defects in SPTA1 are a cause of rhesus-unlinked
 CC elliptocytosis 2 (EL2) (MIM:130600, 182860, 166900).
 CC Elliptocytosis (also known as ovalocytosis) is a genetically
 CC heterogeneous, autosomal dominant hematologic disorder. It is

Query Match

Best Local Similarity 3.5%; Score 134; DB 1; Length 2418;

Matches 175; Conservative 142; Mismatches 307; Indels 280; Gaps 48;

QY 13 GWELSMPEKMSNTNW---VDITQD---PFEACREIKLGLLHDKLGLFEMSAIEM 65
 Db 1572 GNEEAMKEQLKQLEKHEWDHLLERTNDKGLKLEASGRQRFNTSIRDFEFWLSAEETLLAM 1631
 QY 66 MDPKMDAGMIGN-----QVNRKVLNFSQAIKDGTIKIKDLTLPGLIGIMDTFCCLITW 119
 Db 1632 KDQARDLASAGNLKKHQLLEREMLARDELAKDLNTLAEDL----- 1672
 QY 120 LEGHSLAQTVFTCLYIHNPDFI---EDPAMKAFALGILKICIDIAREKVNKA-AVF----- 170
 Db 1673 -----LSSGTF-----NVDQIVKKKNVKNKF-LNVQELAAAHKLEKVALQFFQD 1720
 QY 171 --EEEDF-----QSNTYGFKMANSTDL-----RVTGMLKQVED-----DMQRRVK 209
 Db 1721 LDDERSWIEEKLIRVSSQDYG-RDLQGVQNLLKHKRLEGELVAHEPAIQNVLDNAEKLK 1779
 QY 210 STRSQGGERDEVEL-----EHQCCIAVFSRVKFTVLLTV-----LIAFTKETSAVA 259
 Db 1780 D-KAAVQGE---EIQLRLAQFVEHWEKLEKAKGLKLEESLEYLQFMQNAEEEMIN 1835
 QY 260 EAQKLWVQ--RADLLSAL-----HNSLF-----HGIAQNDDTTKGDHPIMMFFPLVNQ 306
 Db 1836 EKXALAVRGDCGDTLAAQTQSLLMKHEALENDPAVHETRVQNVCAQG-----EDILNK 1887
 QY 307 RLLPPTFPFYAKIIRKREEMVNFABRIDIKTVCEVNLTNLHCILDF-FCFESQSOPCV 365
 Db 1888 VLHEERS-----QNKETSSKIEALNEKTPSLAKAIAAWKLEQLEDDYAFQEFNWKADV 1939

Qy	366	LSRSLQTTFLVDNKKVFGTHL-----MODMKDALRSFVD---PPVLSPKCYLYN- 413
Db	1940	EAMIADKETSLKTNGN--GADLGDFTLLAKQDTLDSLOSFOQERLEPITDLKDLISA 1997
Qy	414	NHQAKDICDSFVTHCVRFPSLIIQHGHNARQDRKLGHLIEEPATLQ---DEPMTFYF- 469
Db	1998	QHNQSKAIEERYAALLKRWEOQLLEASAVHRQ-----LLEKOLFQKAEDLFEVFAHK 2050
Qy	470	-----NRAEKVDAAL-----HTMLLKQ--EPORQH-----LAC 495
Db	2051	ASALNNWCKMEENLSBPFVHCVSLEIRQLQKHEDFLASLARAQADFKCILLEDDQOIKA 2110
Qy	496	LG-----TWLYXNLRMITQYL-----LSCFEL--ELYSMEHYIYIY 530
Db	2111	LGPVSSPYTWLTVEVBERTWKHLSDIIEEREOELQKEARQVKNFQWCQBFQNASTFLQ 2170
Qy	531	WYLSSEFYIYAMLMSTLSRADGSAABERIMBEOQKGRSKKTKKKKKVRLPSREIT----- 585
Db	2171	WILETRAYF-----LDGSLKETGTLESQLEANK-----RKOKETQAMKRQMTKLIVDL 2218
Qy	586	-----MSQAYQ-----NMCAGMFKTWAFDPMQKVRKPKFELDSE 620
Db	2219	GDNLLEDALILDIKYSTIGLAQWDLQVLGLRMOHNLBEOQLQAXDKIKGVSEETLKEPST- 2277
Qy	621	QVRYEHFAPFNSVMTPPVHYLQFKE-MSDLNKYSPPOSPEL----- 663
Db	2278	--IYKH-----FDENLITGLRTH-KEFRSCLRGILNYLPMWVEDEHEPKFEXFLDAVDGPK 2330
Qy	664	YVAASKHFOQAAMI---LENIIMPDPHEVNRILKVAKNFVWMLLAGGHKSKSKVPPEF 719
Db	2331	GYVSLEDY--TAFILDKESENIKSSDIEENAFQALAEKGSYITK-----EDMKQALTEQV 2384
Qy	720	DFA 723
Db	2385	SFCA 2388

RESULT 3			SUB2 HUMAN			STANDARD; PRT; 733 AA.			SUB2 HUMAN			STANDARD; PRT; 733 AA.		
ID	SAB2	HUMAN	Q2UPW6;	Q2UPW6;	Q2UPW6;	Q2UPW6;	Q2UPW6;	Q2UPW6;	Q2UPW6;	Q2UPW6;	Q2UPW6;	Q2UPW6;	Q2UPW6;	
DT	16-OCT-2001	(Rel. 40, Created)	16-OCT-2001	(Rel. 40, Last sequence update)	16-OCT-2001	(Rel. 40, Last sequence update)	16-OCT-2001	(Rel. 40, Last sequence update)	16-OCT-2001	(Rel. 40, Last sequence update)	16-OCT-2001	(Rel. 40, Last sequence update)	16-OCT-2001	
DT	13-MAR-2004	(Rel. 43, Last annotation update)	13-MAR-2004	(Rel. 43, Last annotation update)	13-MAR-2004	(Rel. 43, Last annotation update)	13-MAR-2004	(Rel. 43, Last annotation update)	13-MAR-2004	(Rel. 43, Last annotation update)	13-MAR-2004	(Rel. 43, Last annotation update)	13-MAR-2004	
DE	DNA-binding protein SATB2 (Special AT-rich sequence-binding protein 2).													
DE	SATB2 OR KIAA1034.													
GN	Homo sapiens (Human).													
OS	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;													
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.													
OX	NCBI_TaxID=9606;													
RN	{1}													
RP	SEQUENCE FROM N.A.													
RC	TISSUE=Brain;													
RX	MEDLINE=99397452; PubMed=10470851;													
RA	Kikuno R., Nagase T., Ishikawa K.-I., Hirose M., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;													
RT	"Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";													
RL	DNA Res. 6:197-205(1999).													
RP	{2}													
RX	MEDLINE=22855719; PubMed=12915443;													
RA	FitzPatrick D.R., Carr I.M., McLaren L., Leek J.P., Wightman P., Williamson K., Gautier P., McGill N., Hayward C., Firth H., Markham A.F., Pantes J.A., Bonthron D.T.;													
RT	"Identification of SATB2 as the cleft palate gene on 2q32-q33.";													
RL	Hum. Mol. Genet. 12:2491-2501(2003).													
CC	!- FUNCTION: May play an important role in palate formation.													
CC	!- SUBCELLULAR LOCATION: Nuclear (by similarity).													
CC	!- TISSUE SPECIFICITY: High expression in adult brain, moderate expression in fetal brain, and weak expression in adult liver,													

Db 525 ENPSPNRTL-----WENLC-----TIRRF-----LNLPOHERDVIYEESRHHISRMQ 569
 Qy 632 NSWMTTP-PVHYLQ-----FKMSDLNKYSPPQSPPELYVAASKHFQOAKMILENIP-- 682
 Db 570 HVVQLPPEPVQVLRHQSQAPAKESSPPREBAPPPPTTSCAKKPRRTKISLEALGIL 629
 Qy 683 -----NPDHEVNRIL--KVAKPNFVVMK 703
 Db 630 QSFTHDVGLYDQEAHITLSAQLDLPKHTIHK 661

RESULT 4
 SAB2_MOUSE
 ID SAB2_MOUSE STANDARD; PRT; 733 AA.
 AC Q8VI24;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE DNA-binding protein SATB2 (Special AT-rich sequence-binding protein
 2).
 DE SATB2 OR KIAA1034.
 GN Mus musculus (Mouse).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.
 RX MEDLINE=22855719; PubMed=12915443;
 RA FitzPatrick D.R., Carr I.M., McLaren L., Leek J.P., Wightman P.,
 RA Williamson K., Gautier P., McGill N., Hayward C., Firth H.,
 RA Markham A.F., Fantes J.A., Bonthron D.T.;
 RT "Identification of SATB2 as the cleft palate gene on 2q32-q33";
 RL Hum. Mol. Genet. 12:2491-2501(2003).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Embryonic tail;
 RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S.,
 RA Saga Y., Nagase T., Ohara O., Koga H.;
 RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
 RT III. The complete nucleotide sequences of 500 mouse KIAA-homologous
 RT cDNAs identified by screening of terminal sequences of cDNA clones
 RT randomly sampled from size-fractionated libraries.";
 RL DNA Res. 10:167-180(2003).
 CC -!- FUNCTION: May play an important role in palate formation.
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q8VI24-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q8VI24-2; Sequence=VSP_008967;
 CC Note=No experimental confirmation available;
 CC -!- DEVELOPMENTAL STAGE: Expression first detected at 10.5 dpc in the
 CC maxillary component of the first pharyngeal arch and the lateral
 CC aspect of the frontonasal process in the regions that will
 CC subsequently fuse to form the primary palate. At 11 - 11.5 dpc,
 CC the expression pattern demarcates the region of the medial aspect
 CC of the maxillary process within the primitive oral cavity, which
 CC will form the palate shelf. By 12.5 dpc, symmetrical expression is
 CC seen in the medial edges of the developing palate shelves and this
 CC continues until 13.5 dpc when the strongest expression is in the
 CC mesenchyme underlying the medial edge epithelia. By the time of
 CC palatal shelf fusion at 14.5 dpc the expression is dramatically
 CC down-regulated. No expression detected elsewhere in the embryo at
 CC any stage examined.
 CC -!- SIMILARITY: Belongs to the CUT homeobox family.
 CC -!- SIMILARITY: Contains 2 CUT domains.
 CC -!- SIMILARITY: Contains 1 homeobox domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF319623; AAL37172.1; -;
 DR EMBL; AK129270; BAC98080.1; ALT_INIT.
 DR InterPro; IPR007108; Cut_homeo.
 DR InterPro; IPR003350; Hmoec CUT.
 DR InterPro; IPR001356; Homeobox.
 DR Pfam; PF02376; CUT; 2.
 DR Pfam; PF00046; homeobox; 1.
 DR SMART; SMO0389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX_1; FALSE_NEG.
 DR PROSITE; PS50071; HOMEBOX_2; 1.
 KW Developmental protein; Homeobox; DNA-binding; Nuclear protein; Repeat;
 KW Alternative splicing.
 FT DNA_BIND 350 437 CUT 1.
 FT DNA_BIND 473 560 CUT 2.
 FT DNA_BIND 615 674 HOMEBOX.
 FT VARSPLIC 58 116 Missing (in isoform 2).
 FT /FTid=VSP_008967.
 SQ SEQUENCE 733 AA; 82559 MW; 153CFD1CC3491F25 CRC64;
 Query Match 3.3%; Score 127; DB 1; Length 733;
 Best Local Similarity 20.9%; Pred. No. 0.14;
 Matches 132; Conservative 91; Mismatches 227; Indels 182; Gaps 31;
 Qy 217 EERDEVEL-----EHOQCLAVFSRVKTRVLLTLLVLAFTKETSVAEAKLM----- 265
 Db 67 EQLDGSLEYNREHEAEFLVRKDLVLSQLVETALLALGYSHSSA-AQAQGIILGRWP 125
 Qy 266 -----VQAADLLSALHNSLHHGIGIAQNDDTKGDHP-----IMWGFPL-- 303
 Db 126 LPLSVTDAPDATVADMLQDVYHVVTLLKIQSCSKLEDLPABQWNHATRNALKELLE 185
 Qy 304 VNORLLPTFPFYAKIKKEEMVNYFA-----RLDRIKTV-CEVNLTLN--H 349
 Db 186 MNQSTLKEACPLSQSMISSIVNVTYANVSATKQCFGRWYKVKYKIKVERVERENISDY 245
 Qy 350 CILDFCFEFSQSPCVLSR-----SLTQTFVLDNKVFGTHLMQDMYDALSFDVDPV 404
 Db 246 CVLG-----QRPMLPNNQLASLGTNEQSPHSIHSTPIRNQV-PALQPI MSPGL 297
 Qy 405 LSPK-----CVLYNN-----HOAKDCID-SEVTH-----CVR--PFCSL 435
 Db 298 LSPQLSPQIVRQOIAMAHLINQOIAVSRLLAHQHPQAINQQLNPPPIPAVKPPTNS 357
 Qy 436 IQIHGHNRAQRDKLGHILEEFATLQDEFTFYFNRAEKVDAAHMLLKQBPQHLAC 495
 Db 358 VEVSPIYQVVRDE-----LKRASVSQAVFARVFNRTQ---GLLSEILRKEEDPR-- 405
 Qy 496 LGTWLYNHLRIMIQILSGFELE---LY-----SMHEYIYIYWLS- 535
 Db 406 TASQSLLVNLRAM-QNFLNLPEVERDRYQDERSRMNPVMSVSSASSSSSPSTPOAK 464
 Qy 536 -----FLYAWLMSTLSRADGSQMAERIMEEQKGSKKTKK 574
 Db 465 TSTPTDIPKVDGANVNITAIYDEIQEKKRAKVSQALFAKVANKSQGWLCLELRWK 524
 Qy 575 KKVRELSREITMSQAYQNMCMGFMKTMVAFMDGKVRKPKFELD---SEQVRYEHRFAPF 631
 Db 525 ENPSPNRTL-----WENLC-----TIRRF-----LNLPOHERDVIYEESRHHISRMQ 569
 Qy 632 NSWMTTP-PVHYLQ-----FKMSDLNKYSPPQSPPELYVAASKHFQOAKMILENIP-- 682
 Db 570 HVVQLPPEPVQVLRHQSQAPAKESSPPREBAPPPPTTSCAKKPRRTKISLEALGIL 629
 Qy 683 -----NPDHEVNRIL--KVAKPNFVVMK 703
 Db 630 QSFTHDVGLYDQEAHITLSAQLDLPKHTIHK 661
 RESULT 5

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Precursor of the egg-yolk proteins that are sources of
 CC nutrients during embryonic development (Potential).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Contains 1 VWFD domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AC024137; AAK09075.1; --
 DR WormPep; F59D8.1; CE20900.
 DR InterPro; IPR001747; Lipid transprt_N.
 DR InterPro; IPR001846; VWF D.
 DR Pfam; PF01347; Vitellogenin_N; 1.
 DR Pfam; PF00094; vwd; 1.
 DR SMART; SM00638; LPD_N; 1.
 DR SMART; SM00216; vwd; 1.
 DR Storage protein; Multigene family; Signal.
 KW SIGNAL 1 15
 FT CHAIN 16 1603
 FT DOMAIN 1308 1455
 FT SEQUENCE 1603 AA; 186529 MW; BCA02762477D37DE CRC64;
 SQ
 Query Match 3.3%; Score 125.5; DB 1; Length 1603;
 Best Local Similarity 20.0%; Pred. No. 0.52;
 Matches 117; Conservative 89; Mismatches 234; Indels 145; Gaps 25;
 234 FSRVKTFRVLLVLAFAFKKTSVAEAKLMQVQADLSLHGIQANDTTKGD 293
 706 FQQFEQVILKTL-----EKLISLKGKSDLSLRSVQSGIOMLQEIIVK-- 749
 294 HPIMGFEPLVNRQLPPTFPFYAKIKREEMVNFARLIDRIKTVCEVNV----- 344
 750 ---KNIRPRVQ---TDSQNAHAFYLYKEMDVIPLID-MEIDNVVEKVRNGEPD 802
 345 -----LTN-----LHCILDFCFEFSQSPVLSRLSLL-----QTFVLVNDKKV----- 382
 803 IKSLLTFLTNDKSFELHRL-FFYZAERRIPTTIGMPLTISGKMPITLSINGKVSIELEK 861
 383 FGTHLAQDMVXDALRSFVDP-----PVLSPKCYLYNNHQAKDCIDSFVTHCVRPSCSLIQ 437
 862 LGARLVLDIVPTVAHTVEMFWFVFI-----EGQVKLSQSLARLHTPLRFESTIVE 912
 438 IHGHNARQRDKLGHILEFAFLQDEFMTF-----YFNRAEKVDA 478
 913 L-----KKNTL-BITHKFVVPENKKTIVSHTRPVAFIRVPKQDSYVTEKTI 964
 479 LHTMLLKQEPQOHACLGTVVLYHNLRMIO-YLLSGFEELYSMHEYYIYWLVS--- 534
 965 SOYQMSTIEDROYET-----FGLRINAGNVLSQWTLPMWLMTQDPEFTLENK 1016
 535 --EFLYAWLMTLSRADQSQAEEIMEQOGRSKTKTKKKKVRPLSREITMCAQVN 592
 1017 PVFEARTVITGNLEKTDLSIEIKFDKIFEKFDLENNESENRRQYFHKMIREQSGFGKN 1076
 593 MCA-----GMFKTMAFMDGKVRKPKFELDSQVRYEHRAPEN-----SVMIPP 638
 1077 LITLKLKAPQMYNWTETRTVCDKWRMCKVNDARRSPMEHNEKWTURTELLAARQ 1136
 639 PVHYIQFKEMSD-----IN-KYSPPPQSPELYVAASKHFQQAQKMLLENIPN-----PDH 686

Db 1137 PSSLRQLREQPHREYQLALNAKWSSKSEITFNAQLQEQSTQKKFLRNIEREYKGIPEY 1196
 QY 687 EVNRILKVAKPNFVVMKLIAGHKFKESKVPPEDFSAHKYFPVVK 731
 Db 1197 EL--LIKAARLNQVNV-----VSEYKLTPESEYTFSRIFDLIK 1232
 RESULT 7
 VIT4 CAEEL
 ID VIT4 CAEEL STANDARD; PRT; 1603 AA.
 AC P18947; Q9BPP3;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Vitellogenin 4 precursor.
 GN VIT-4 OR F59D8.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1-282 FROM N.A.
 RA Blumenthal T., Spieth J., Zucker E.;
 RL Submitted (NOV-1986) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-71 FROM N.A.
 RX MEDLINE=85269643; PubMed=4022780;
 RA Spieth J., Denison K., Kirtland S., Cane J., Blumenthal T.;
 RT "The C. elegans vitellogenin genes: short sequence repeats in the
 RL Nucleic Acids Res. 13:5283-5295 (1985).
 CC -!- FUNCTION: Precursor of the egg-yolk proteins that are sources of
 CC nutrients during embryonic development (Potential).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: Synthesized in Caenorhabditis only by 32 cells
 CC building the intestine of adult hermaphroditic individuals; they
 CC are cotranslationally secreted into the body cavity and
 CC subsequently taken up by the gonad.
 CC -!- SIMILARITY: Contains 1 VWFD domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AC024137; AAK09074.1; --
 DR EMBL; M11498; AAA28163.1; --
 DR EMBL; X02754; CAA26531.1; --
 DR PIR; A43084; A43084.
 DR WormPep; F59D8.2; CE26817.
 DR InterPro; IPR001747; Lipid transprt_N.
 DR InterPro; IPR001846; VWF D.
 DR Pfam; PF01347; Vitellogenin_N; 1.
 DR SMART; SM00638; LPD_N; 1.
 DR SMART; SM00216; vwd; 1.
 DR Storage protein; Multigene family; Signal.
 KW SIGNAL 1 15
 FT CHAIN 16 1603
 FT DOMAIN 1308 1455
 FT SEQUENCE 1603 AA; 186307 MW; E303170325BC99BB CRC64;
 FT SIGNAL 1 15
 FT VITELLOGENIN 4.
 FT DOMAIN 1308 1455
 FT VWF D.
 FT Y -> V (IN REF. 3).
 FT L -> V (IN REF. 2).
 FT EVAYT -> RSLRH (IN REF. 2).
 FT T -> S (IN REF. 2).
 FT SEQUENCE 1603 AA; 186307 MW; E303170325BC99BB CRC64;

SEQUENCE FROM N.A.
 RP TISSUE=Muscle;
 RC MEDLINE=89210800; PubMed=3072195;
 RA Lemaire C., Heilig R., Mandel J.L.;
 FT "The chicken dystrophin cDNA: striking conservation of the C-terminal
 RT coding and 3' untranslated regions between man and chicken.";
 RL EMBO J. 7:4157-4162(1988).
 CC -!- FUNCTION: May play a role in anchoring the cytoskeleton to the
 CC plasma membrane.
 CC -!- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY
 CC ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN,
 CC ABP-120, ABP-180, OR BETA-FODRIN).
 CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
 CC -!- SIMILARITY: Contains 22 spectrin repeats.
 CC -!- SIMILARITY: Contains 1 WW domain.
 CC -!- SIMILARITY: Contains 1 ZZ-type zinc finger.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X13369; CRA31746.1; -.
 DR PIR; S02041; S02041.
 DR HSP; P46939; 1BHD.
 DR InterPro; IPR001589; Actbind actinin.
 DR InterPro; IPR001715; Calponin-like.
 DR InterPro; IPR002017; Spectrin.
 DR InterPro; IPR001202; WW Rsp5_WMP.
 DR InterPro; IPR000433; Znf_ZZ.
 DR Pfam; PF00307; CH; 2.
 DR Pfam; PF00435; spectrin; 21.
 DR Pfam; PF00397; WW; 1.
 DR Pfam; PF00569; ZZ; 1.
 DR SMART; SMO0033; CH; 2.
 DR SMART; SMO0150; SPEC; 21.
 DR SMART; SMO0456; WW; 1.
 DR SMART; SMO0291; Znf_ZZ; 1.
 DR PROSITE; PS00019; ACTININ_1; 1.
 DR PROSITE; PS00020; ACTININ_2; 1.
 DR PROSITE; PS00021; CH; 2.
 DR PROSITE; PS00020; WW DOMAIN 1; 1.
 DR PROSITE; PS01357; ZF_ZZ_1; 1.
 DR PROSITE; PS0135; ZF_ZZ_2; 1.
 KW Structural protein; Actin-binding; Calcium-binding; Cytoskeleton;
 KW Repeat; Zinc-finger.
 FT DOMAIN 1 244
 FT DOMAIN 19 123
 FT DOMAIN 138 241
 FT REPEAT 341 449
 FT REPEAT 450 558
 FT REPEAT 561 669
 FT REPEAT 721 830
 FT REPEAT 832 936
 FT REPEAT 945 1047
 FT REPEAT 1050 1156
 FT REPEAT 1159 1265
 FT REPEAT 1268 1369
 FT REPEAT 1470 1570
 FT REPEAT 1573 1678
 FT REPEAT 1681 1782
 FT REPEAT 1879 1981
 FT REPEAT 2013 2103
 FT REPEAT 2106 2211
 FT REPEAT 2214 2321
 FT REPEAT 2472 2574
 FT REPEAT 2577 2683
 FT REPEAT 2686 2799
 FT REPEAT 2802 2904

FT REPEAT 2906 2928 SPECTRIN 21.
 FT REPEAT 2931 3037 SPECTRIN 22.
 FT DOMAIN 3052 3085 WW.
 FT ZN_FING 3304 3351 ZZ-TYPE.
 FT VARIANT 1171 1171 MISSING.
 FT VARIANT 1869 1869 Q -> H.
 FT VARIANT 1885 1885 K -> R.
 SQ SEQUENCE 3660 AA; 422874 MW; 85493DAF6D5B6D4A CRC64;
 Query Match 3.1%; Score 117.5; DB 1; Length 3660;
 Best Local Similarity 17.7%; Pred. No. 6;
 Matches 132; Conservative 113; Mismatches 229; Indels 273; Gaps 35;
 QY 13 GWELSMPEKMEKSNNTNWVDIT----- 33
 DB 1646 GKSIVDEKLSLNSNIAVTISRAEENLNLMEYKMEAEQKVAVTTWYRAEILLD 1705
 QY 34 -----QDFEACREIKLGEI--LHKLFGLFRAMSAIENM-----DPKMDAG 73
 DB 1706 ESDKQKQKQKEETLKRLLK-AELNDMHPKVDSDVRD--QAVDLMTNRGDHCRKVIIEPKL--- 1759
 QY 74 MIGNQVNRKVLNFOAIDKGITIKIKDLTLPELIGIMDTFCCLITWLBGHSIAQTFTCL 133
 DB 1760 ---SELNHRFAAISQRIKSG----- 1776
 QY 134 YIHPDFTEDPAMKAFALGILKICDIAREKVNKAAPFEEDF-----QSMTY 180
 DB 1777 ---KPIPLKEQDFDIQKLEPLEVEIQGVNLKEEDFNKDMSEDDSTVKELLQR 1832
 QY 181 GFKMANSVTDLRVTGMLKQVEDDMORRVKSTRSQGEERDPEVELEHQOCLAVSRVKT 240
 DB 1833 GDTLQKRIITDERKEEIKIKQQLLQTHNALKDLRSRRKKALEISHOW----YQYKQA 1888
 QY 241 RVLLTVLIAFTFKETSAV--AAQOKLVQAAADL-----LSAHSNLHHGIAQNDTTKG 292
 DB 1889 DDLMTWLDIEKKLASUPDHKDEQKLEIGELEKKKEDLNANVR-----QAEKLSKDG 1942
 QY 293 DHPIMMGFEPLVNQRLLPPTFPYAKIKREEMVNYFARLIDRIKTVCEVVNLTNLHCIL 352
 DB 1943 -----AAKAVEPTLVQLSK--RWEDFSKTAQF-----RLNYAQIQITVL 1980
 QY 353 DFCFEPSEQPCVLSRL--LQTTFLVNDKKNKVFVGHLMQDM--VKDALRSFVDPVPLSPKC 409
 DB 1981 -----EDTTFVMTESMTVETVPSTYLAELIQLLQALSEVERLNS---PVL----- 2025
 QY 410 YLYNNHOAKCIDSF-VTHCVRPF--CSLIQIHGHNRRARQDKLGHIL---EBPATIQD 462
 DB 2026 -----QAKCEDLLKQEECLNKKDC--LGRLOGHIDIIHKKTPALQASATPRETANIQD 2078
 QY 463 EFMFTYFNRAEKVDAALHTMLLKQEPQORQHLACIGTWLVYHNLIRIMIQYLLSGFELELYS 522
 DB 2079 K-LTQLNSQWEKVNMYRDRQARFDSKE-----KWRLFH-----CEMKS 2117
 QY 523 MHEYYYIYWLSEFLYAWLMT---LSRA-----DGSQMAEERIMEEQKRSKKTKK 574
 DB 2118 FNE-----WLTETEELKSLRAQIEAGDVGHVTKQFLQELQDQIGIGQQTVVK 2163
 QY 575 KKVRLPSREITMSQAYQNMCMAGMFKTVAFMDCKVRKPKF-----ELDSQVRYEH 626
 DB 2164 -----TINVIGEEIIEQSSNA-----DANVLKEQLGNLNRWQEICRQLVEKKR 2207
 QY 627 RFAPFNSVMTPPPVHYIQFKEMSDLNK 653
 DB 2208 RIEEKNILS-----EFQE--DLNK 2225
 RESULT 11
 BAC1 HUMAN
 ID BAC1 HUMAN STANDARD; PRT; 736 AA.
 AC O14867; O43285;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)


```
Db      718 YEIS---KLKKTASFIPTKESLTFDFQCKKELQMLRKESEISN----- 763
Qy      63 IEMDPKMDAGNIGVNRKVL-----NFEQAKOGTIKIDLTPELIGIMDTOPCCILIT 118
Db      764 ---ENKMDFSKGGYKAKIKELNNLERLSDLQSKIQE-----LESIRSKDSOLK 813
Qy      119 WLEGHSLAQVTCTLIYHNPDPFIEDPAMKAFALGILKICDIAREKVNKAAYFE----- 171
Db      814 WAQ-----NTIDTEMKMSL-----LTLSNKETIEKLSSEIE 848
Qy      172 --EEDFQSMYGFPMKANSVTDL-----RVTKMLKDVEDDMQR----- 206
Db      849 NLDKELRKTFOYKFLDQNSDASTLEPTLRKELEQIQVLQKDNASQIQAYBEIISNENA 908
Qy      207 --RVKSTRSGEERDPVELEHQCLIA----- 232
Db      909 LIEKLNELAKTKENYDAKIELEKKEKWARBEEDLSRLGELGSEIRALQPKLKEGALHFVQQ 968
Qy      233 -----VFSRVKFTRVLLTVLIAFTKETSAAVAQKLMVQAAADLLSAIHNSLHHG 282
Db      969 SEKLRNEVERIQKLEKIEKSTIVQLCKKEMS---QYQSTMKENKOLSELVIRLEKDA 1025
Qy      283 IQANDTTKGDPHIMGFEPLVNRQLPPTFPFYAKIITKREEMVNYFARLIDRIKTV--C 340
Db      1026 ADCQAEKTKSSLYSA-----QDLKDKHERKW---MEEKADYERELISNIEQTESL 1074
Qy      341 EVMNLNLEACILDFPCEFSBQPCVL-----SRSLQTTFLVDVNNKVPFGTHLMQDM 391
Db      1075 RVENSVLIEKVDVDTAANNQDKDLKLVLSFNLRHERNSLETKLTCKRELAFVKQKND 1134
Qy      392 VKDALRSVPDPVLPSPKCYLYNNHQAQDC---IDSF--VTHCVRPFCSLIQIHGHNAR 445
Db      1135 LEKINDIQRQTISEKY-----QCSAVIIDEFKDITKEV---TQVNLKENNAI 1182
Qy      446 QRDKLGHILEE---FATLODEBMTFFYFNRAEKVDAAHLTMLKQEPQORHLACIGTWVL 501
Db      1183 LQKSLKNVTEKNREIYKQND-----RQBEI--SRLQRLDIQTKEQ-----VS 1223
Qy      502 YHNLRIIMQYLLSGPELELYSMHEYY-----IYWLSSEFLYAWLMSTLSRA 548
Db      1224 INSNKILV-----YESEMEQCKQYDLSQOQKDAQKDIKLTNEI--SDLKGLKLSA 1275
Qy      549 DGSQMAEBRIMEEQKGRSSKTKKKKVRPLSREITMSQAYQNNCAG--MFKTMTVAEDM 606
Db      1276 ENANADLENKFNRLKQAEKLDASKQQAALTNELNELKAIKKDLQDLHPENAKVIDL 1335
Qy      607 DGKVRKPKFELDSEQVRYEHRFAPFNSVMTPPVHYLQFKEMSDLNKYSPPQSPELYV- 665
Db      1336 DTKLK--AHELQSEDSVRDHEKDTYRTLM-----EETESLKK-----ELQIF 1375
Qy      666 ----AASKHFQQAAMLENIENPDDEVNRIL 692
Db      1376 KTNSSSDAFEXLKV-----NMEXEKDRII 1400
RESULT 14
SPOF_SCHPO STANDARD; PRT; 1957 AA.
AC Q10411; Q9USE9;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Sporulation-specific protein 15.
GN SPO15 OR SPAC1F3.06C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN
RP SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
RX MEDLINE=20107136; PubMed=10639340;
```

```
RA Ikemoto S., Nakamura T., Kubo M., Shimoda C.;
RT "S. pombe sporulation-specific coiled-coil protein Spo15p is localized
RL to the spindle pole body and essential for its modification.";
RN J. Cell Sci. 113:545-554 (2000).
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weetjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsberg S.L.,
RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880 (2002).
RP SEQUENCE OF 705-871 FROM N.A.
RC STRAIN=968 h90;
RX MEDLINE=2023868; PubMed=10759889;
RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,
RA Hiraoka Y.;
RT "Large-scale screening of intracellular protein localization in living
RL fission yeast cells by the use of a GFP-fusion genomic DNA library.";
CC Genes Cells 5:169-190 (2000).
CC -!- FUNCTION: Has a role in the initiation of spore membrane
CC formation.
CC -!- SUBUNIT: Monomer.
CC -!- SUBCELLULAR LOCATION: Spindle pole body.
CC -!- SIMILARITY: Belongs to the MPC70 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z70690; CAA94624.1; -.
CC EMBL; AB027811; BAA87115.1; -.
DR PIR; T38077; T38077.
DR GeneDB SPombe; SPAC1F3.06c; -.
KW Sporulation; Coiled coil.
FT DOMAIN 199 785 COILED COIL (POTENTIAL).
FT DOMAIN 804 1235 COILED COIL (POTENTIAL).
FT DOMAIN 1320 1471 COILED COIL (POTENTIAL).
FT DOMAIN 1481 1723 COILED COIL (POTENTIAL).
SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;
Query Match 2.9%; Score 113.5; DB 1; Length 1957;
Best Local Similarity 21.1%; Pred. No. 5.1;
Matches 113; Conservative 86; Mismatches 215; Indels 121; Gaps 24;
Qy 9 DDDSGWELSMPEKESKNTNWDITQDPFEACRELKLGEL--LHDKLFGLEAMSALW 66
```

Db 1449 DTSRKSSLMKIESINSSLD--KSPELASAVEKLGALQKJHSESLSMENIKS-OLQ 1505
Qy 67 DPK----MDAGMI---GNQVNRKVLNFEQAKDGTIKIKDLT-----LPELIGMDTCFC 114
Db 1506 EAKKIQVDESTIQELDHETITASKNNYEGKLNKDSIIRDLSENIEQNNLLAEKSAVK 1565
Qy 115 CLITWLEHSLAQVFTCLYIHNPDFIEDPAKAFALGILKICDIAREKVNKAAPFER-- 172
Db 1566 RLSTKESEILQFNSRLADLEYHKSOVES-----ELGRSKKLASTTELO 1611
Qy 173 ----EDFOSMTYGYKMANSVTDLRVGMLKVEDDMQRRVKSTRSQGEERDPVELEHQ 228
Db 1612 LAENERLSLATRMLDLQNVKD-----LSNIKDSLSBDLRLASLE-----DSVASLQ-K 1660
Qy 229 QCLAVFSRVKTRVLLTVLIAFT---KKESTAVAEAKLMVQAADLLSAIHNSLHHGLOA 285
Db 1661 ECKIKNSVTSLQDLVTSVQARNAELEDEVSVDKIRRRDRCHLSGKLUKLSQLE 1720
Qy 286 QNDT---TKGHPIMMGF--BPLVNQ-RLLPPTFPFYAKIIRBEMVNYFARLIDRIKT 339
Db 1721 QHETFFRAEQRMQLGFLKETVKVKQKLLKLNLRQQLIPRSILVYESYIRDIEKEI 1780
Qy 340 CEVNTNLNHCILDFCFSPSQSPCVLSRSLQTTFLVDNKKVFGTHLMQDMVKDALRSF 399
Db 1781 --IVLQRLNGI-----ELSQQLP-----KGYFG----- 1802
Qy 400 VDPFVLPSPKVLNNHQAQ-CDIOSFVTHCVRPFCSLIQHGNRQORDKLGHLIERFA 458
Db 1803 -----YFFKTRNVEMLDSFKQOVAK-----IQFLAG---AEFIVRFKEDLEKCA 1845
Qy 459 TLODFMTFFYNRAEKVD-----AALHTMLLKQEPQROHLAGLGTWVLYHNLR 507
Db 1846 AEEKEKQATPDYSEKVENLGKSI EALYPALNREISFEKSLALS KS--AYENLLV 1898

RESULT 15

TELE YEAST
ID TELE YEAST STANDARD; PRT; 2787 AA.
AC P38110;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Telomere length regulation protein TEL1.
GN TEL1 OR YBL088C OR YBL0706.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=96076635; PubMed=7502586;
RA Obermaier B., Gassenhuber J., Piravandi E., Domdey H.;
RT "Sequence analysis of a 78.6 kb segment of the left end of
RT Saccharomyces cerevisiae chromosome II.";
RL yeast 11:1103-1112(1995).
RN [2]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RA Greenwell P.S., Krommal S.L., Porter S.E., Gassenhuber J.,
RA Obermaier B., Petes T.D.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Putative phosphatidylinositol kinase involved in
CC controlling telomere length.
CC -!- SIMILARITY: Belongs to the PI3/P14-kinase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).

CC -----CAA56016.1; -
DR EMBL; X79489; CAA56016.1; -
DR EMBL; Z35849; CAA84909.1; -
DR EMBL; U31331; AAA69802.1; -
DR PIR; S45416; S45416.
DR Germline; 138523; -
DR SGD; S0000184; TEL1.
DR GO; GO:0005634; C:nucleus; IC.
DR GO; GO:0007004; P:telomerase-dependent telomere maintenance; IMP.
DR InterPro; IPR003151; PAT.
DR InterPro; IPR003152; PAT.
DR InterPro; IPR000403; P13_P14_kinase.
DR Pfam; PF02259; PAT; 1.
DR Pfam; PF02260; PAT; 1.
DR Pfam; PF00454; P13_P14_kinase; 1.
DR SMART; SM00146; P13Kc; 1.
DR PROSITE; PS00915; P13_4_KINASE_1; 1.
DR PROSITE; PS00916; P13_4_KINASE_2; 1.
DR PROSITE; PS02090; P13_4_KINASE_3; 1.
KW Transferrase; Kinase.
FT DOMAIN 2461 2787 P13K/P14K.
FT CONFLICT 1190 1190 Y -> F (IN REF. 2).
SQ SEQUENCE 2787 AA; 321663 MW; 439B6E189E39499B CRC64;
Query Match 2.9%; Score 113; DB 1; Length 2787;
Best Local Similarity 22.4%; Pred. No. 9;
Matches 67; Conservative 50; Mismatches 96; Indels 86; Gaps 17;
Qy 323 BEMVNYFARLIDRI-KTVCEV-VNLTNLHCILDFCFSESPCVLSRSLQTTFLVDNK 380
Db 1346 DEVITIFSSLNADKNTFEIPEPLNLFCKIFTYLRNKQ---LSPFQQAIIKLEHR 1401
Qy 381 -----KVFQTHLMQDMVKDALRSFVDPVLPSPKVLNNHQ---AKDC-IDSEFV--- 425
Db 1402 DLIIKIKWKYFLDAIFGNVQDD-----IYENTELLDSADCGVDDVLVS 1446
Qy 426 ---THCVRPF-----CSLIQIHGHNRRQORDKLGHLIEEFATLODFMTFFYNRAEKVDA 477
Db 1447 LLFSVARRPVASKIGCSLSKAAAINILKH-----HVPKEY--LSKNFKLWF-----A 1491
Qy 478 ALHTMLLKQEPQROHLAGLGTWVLYHNLR-----MIQYLLSGP--ELELYSMHEY 526
Db 1492 ALSRRILQEQVERERSTNFNNEVHLKNFEMVFRHPEQPHMYQRISTFNKEAELDYDSTE 1551
Qy 527 YVIVYWLSEFLYALMS-TLSRADGSQMAEERIMEEQKGRSSKTKKKKKVRLPLSREI 584
Db 1552 FFI-----SECILTVLVGYSIGNSESEFCFRDNINNE-----NKDKVAPLDKDV 1595

Search completed: August 10, 2004, 21:50:35

Job time : 17 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 21:48:07 ; Search time 43 Seconds
(without alignments)
5378.485 Million cell updates/sec

Title: US-10-001-857-145

Perfect score: 3849

Sequence: 1 MVMKASVDDDDSGWELSMPE.....KVPPEDFSAHKRPVVKLV 733

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_page:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertibrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3684	95.7	725	11 Q8K3H2	Q8k3h2 mus musculus
2	3648	94.8	725	11 Q9J101	Q9j101 rattus norv
3	3331.5	86.6	724	13 Q7T322	Q7t322 brachydanio
4	2670	69.4	526	4 Q9H703	Q9n703 homo sapien
5	1326	34.5	784	5 Q9W1A2	Q9w1a2 drosophila
6	1116	29.0	211	4 Q9H631	Q9h631 homo sapien
7	1017.5	26.4	708	5 Q8T3X5	Q8t3x5 drosophila
8	772	20.1	150	11 Q8BYJ9	Q8byj9 mus musculus
9	584	15.2	695	10 Q9SKH7	Q9skh7 arabidopsis
10	560.5	14.6	799	5 Q17003	Q17003 caenorhabdi
11	377.5	9.8	908	3 Q8WZV5	Q8wzv5 neurospora
12	252.5	6.6	708	3 Q9USY3	Q9usy3 schizosacch
13	190	4.9	41	11 Q8K3H1	Q8k3h1 mus musculus
14	150	3.9	1218	5 Q81338	Q81338 plasmodium
15	139.5	3.6	964	13 Q8JGR7	Q8jgr7 brachydanio
16	131	3.4	2965	5 Q8WRS5	Q8wrs5 plasmodium

17	130.5	3.4	893	5 Q8I3I6	Q8i3i6 plasmodium
18	129	3.4	2976	5 Q8WRS4	Q8wrs4 plasmodium
19	127	3.3	733	11 Q8V124	Q8v124 mus musculus
20	127	3.3	969	16 Q9ZJA3	Q9zja3 helicobacte
21	126	3.3	2770	5 Q7YUE9	Q7yue9 plasmodium
22	126	3.3	2957	5 Q8WRS6	Q8wrs6 plasmodium
23	123.5	3.2	778	5 Q19642	Q19642 caenorhabdi
24	122.5	3.2	728	11 Q8K454	Q8k454 mus musculus
25	122.5	3.2	1607	5 Q20187	Q20187 caenorhabdi
26	122.5	3.2	2253	13 P70012	P70012 xenopus lae
27	122	3.2	1175	16 Q8XNW6	Q8xnw6 clostridium
28	121	3.1	966	11 Q8C3A9	Q8c3a9 mus musculus
29	121	3.1	966	11 Q7TQK1	Q7tqk1 mus musculus
30	119	3.1	986	4 Q8TBW2	Q8tbw2 homo sapien
31	119	3.1	1196	4 Q94834	Q94834 homo sapien
32	119	3.1	3412	4 Q60316	Q60316 homo sapien
33	118.5	3.1	550	11 Q8BI22	Q8biz2 mus musculus
34	118.5	3.1	1029	10 Q80955	Q80955 arabidopsis
35	118.5	3.1	5303	5 Q9V628	Q9v628 drosophila
36	118	3.1	728	4 Q60561	Q60561 homo sapien
37	118	3.1	793	5 Q815C9	Q815c9 plasmodium
38	116.5	3.0	1045	10 Q94G20	Q94g20 daucus caro
39	116.5	3.0	1641	11 Q88528	Q88528 mus musculus
40	115.5	3.0	550	11 Q8BIL5	Q8bil5 mus musculus
41	115	3.0	721	5 Q27308	Q27308 anopheles a
42	115	3.0	1275	4 Q8TE30	Q8te30 homo sapien
43	115	3.0	1316	10 Q7Y239	Q7y239 oryza sativ
44	114	3.0	546	12 Q98682	Q98682 simian cyto
45	114	3.0	865	6 Q95LL1	Q95ll1 macaca fasc

ALIGNMENTS

RESULT 1

Q8K3H2 ID Q8K3H2 PRELIMINARY; PRT; 725 AA.
AC Q8K3H2;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DE 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Embryonic growth-associated protein EGAP.
GN C030004C14RIK.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Wenzlau J.M., Weisner-Evans M.C.M.;
RT "Mouse Embryonic Growth Associated Protein (EGAP).";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY102701; AAM52342.1; -;
DR MGD; MGI:1925939; C030004C14RIK.
DR InterPro; IPR007244; Mak10.
DR Pfam; PF04112; Mak10; 1.
SQ SEQUENCE 725 AA; 83349 MW; F9C7E798F7B51DFC CRC64;

Query Match	95.7%	Score 3684;	DB 11;	Length 725;
Best Local Similarity	95.9%	Pred. No. 2.6e-300;		
Matches 703;	Conservative 11;	Mismatches 11;	Indels 8;	Gaps 1;
QY	1	MVMKASVDDDDSGWELSMPEKESNTNWVITQDFEACHRELKLGELHDKLFLGFAM	60	
		: : : : : : : : :		
Db	1	MVMKATVDDDSAGMELGVPKMKESSTSWVDITQDFEDACRELKLGELHDKLFLGFAM	60	
QY	61	SAIEMMDPKMDAGMIGNVNRKVLNFEQAIKDGTTIKIKDLTIPELIGIMDTCFCLITWL	120	
		: : : : : : : :		
Db	61	SAIEMMDPKMDAGMIGNVNRKVLNFEQAIKDGTTIKIKDLTIPELIGIMDTCFCLITWL	120	
QY	121	EGHSLAQIVFTCLYTHNPFDFIEDPAMKAFALGILKICIDIAREKVNKAAVFEEEDFQSMTY	180	
		: : : : : : : :		
Db	121	EGHSLAQIVFTCLYTHNPFDFIEDPAMKAFALGILKICIDIAREKVNKAAVFEEEDFQSMTY	180	

```

QY 181 GFKMANSVTDLRVVTGMLKDVEDDMORRVKSTRSQGEERDPEVELEHOOCLAVSRVKFT 240
Db 181 GFKMANSVTDLRVVTGMLKDVEDDMORRVKSTRSQGEERDPEVELEHOOCLAVSRVKFT 240
QY 241 RVLTLVIAFTKKTSAVAEAKLMVQAADLLSAIHSNLSHHGIOAQNDTTKGDHPIMMGF 300
Db 241 RVLTLVIAFTKKTSAVAEAKLMVQAADLLSAIHSNLSHHGIOAQNDTTKGDHPIMMGF 300
QY 301 EPLVNQRLPPTFPFYAKIIKREEMVNYFARLIDRIKTVCEVNVNLTNHLCHILDFCFEPE 360
Db 301 EPLVNQRLPPTFPFYAKIIKREEMVNYFARLIDRIKTVCEVNVNLTNHLCHILDFCFEPE 360
QY 361 QSPCVLSRLSLLQTTFLVDNKKVFGTHLMQDMVKALRSFVPPVLSPKCYLYNNHQKDC 420
Db 361 QSPCVLSRLSLLQTTFLVDNKKVFGTHLMQDMVKALRSFVPPVLSPKCYLYNNHQKDC 420
QY 421 IDSFVTHCVRFPCSLIQIHGHNARQDKLGHILBEEFATLQDEFTFYFNRAEKVDAALH 480
Db 421 IDSFVTHCVRFPCSLIQIHGHNARQDKLGHILBEEFATLQDEFTFYFNRAEKVDAALH 472
QY 481 TMLLKQEPQORHCLAGTWLVYHNLRIIMIQYLLSGFELELYSMHYYIYIYWLSEFLYAW 540
Db 473 TMLLKQEPQORHCLAGTWLVYHNLRIIMIQYLLSGFELELYSMHYYIYIYWLSEFLYAW 532
QY 541 LMSTLSRADGSGMAEERIMEEQOGRSSKTKKKKKVRLPGLREITMSQAYQNMCAGMFKT 600
Db 533 LMSTLSRADGSGMAEERIMEEQOGRSSKTKKKKKVRLPGLREITMSQAYQNMCAGMFKT 592
QY 601 MVAEDMDGKVRKPKFELDSQVRYEHRFAPNSVMTPTPPVHYLPKEMSDLNKYSPPPQS 660
Db 593 MVAEDMDGKVRKPKFELDSQVRYEHRFAPNSVMTPTPPVHYLPKEMSDLNKYSPPPQS 652
QY 661 PELVYAAKSHFQQAQKMLENIPNDPHEVNRILKVAKNPVMVVKLLAGGHKKESKVPPEFD 720
Db 653 PELVYAAKSHFQQAQKMLENIPNDPHEVNRILKVAKNPVMVVKLLAGGHKKESKVPPEFD 712
QY 721 FSAHKYFPVVKLV 733
Db 713 FSVHKYFPVVKLV 725

RESULT 2
Q9JIO1
ID Q9JIO1 PRELIMINARY; PRT; 725 AA.
AC Q9JIO1
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Corneal wound healing related protein.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20314365; PubMed=10855038;
RA Yi X.J., Li X.F., Yu F.S.;
RT "A novel epithelial wound-related gene is abundantly expressed in
RL Curr. Eye Res. 20:430-440(2000).
DR EMBL; AF272892; AAF81791.1; -
DR InterPro; IPR007244; Mak10.
DR Pfam; PF04112; Mak10; 1.
SQ SEQUENCE 725 AA; 83151 MW; ADB38695BDE1CD2D CRC64;

Query Match 94.8%; Score 3648; DB 11; Length 725;
Best Local Similarity 95.0%; Pred. No. 2.7e-297;
Matches 696; Conservative 15; Mismatches 14; Indels 8; Gaps 1;

QY 1 MYMKASVDDDDSGWELSPKEMKSNWNVDITQDFEACRELKGLGELHDLKGLGFLFAM 60
Db 1 MYMKAAVDDDDASGWELNVPKEMKSNWNVDITQDFEACRELKGLGELHDLKGLGFLFAM 60

```

```

QY 61 SAIEWMDPKMDAGMIGNOVNKKVLNFEQAIKDGTIKIKDLTLPELIGIMDTFCCLITWL 120
Db 61 SAIEWMDPKMDAGMIGNOVNKKVLNFEQAIKDGTIKIKDLTLPELIGIMDTFCCLITWL 120
QY 121 EGHSLAQVTFCTLYIHNPDFIEDPAMKAFALGILKICDIAREKNVKAAPFEEDFQSMY 180
Db 121 EGHSLAQVTFCTLYIHNPDFIEDPAMKAFALGILKICDIAREKNVKAAPFEEDFQSMY 180
QY 181 GFKMANSVTDLRVVTGMLKDVEDDMORRVKSTRSQGEERDPEVELEHOOCLAVSRVKFT 240
Db 181 GFKMANSVTDLRVVTGMLKDVEDDMORRVKSTRSQGEERDPEVELEHOOCLAVSRVKFT 240
QY 241 RVLTLVIAFTKKTSAVAEAKLMVQAADLLSAIHSNLSHHGIOAQNDTTKGDHPIMMGF 300
Db 241 RVLTLVIAFTKKTSAVAEAKLMVQAADLLSAIHSNLSHHGIOAQNDTTKGDHPIMMGF 300
QY 301 EPLVNQRLPPTFPFYAKIIKREEMVNYFARLIDRIKTVCEVNVNLTNHLCHILDFCFEPE 360
Db 301 EPLVNQRLPPTFPFYAKIIKREEMVNYFARLIDRIKTVCEVNVNLTNHLCHILDFCFEPE 360
QY 361 QSPCVLSRLSLLQTTFLVDNKKVFGTHLMQDMVKALRSFVPPVLSPKCYLYNNHQKDC 420
Db 361 QSPCVLSRLSLLQTTFLVDNKKVFGTHLMQDMVKALRSFVPPVLSPKCYLYNNHQKDC 420
QY 421 IDSFVTHCVRFPCSLIQIHGHNARQDKLGHILBEEFATLQDEFTFYFNRAEKVDAALH 480
Db 421 IDSFVTHCVRFPCSLIQIHGHNARQDKLGHILBEEFATLQDEFTFYFNRAEKVDAALH 472
QY 481 TMLLKQEPQORHCLAGTWLVYHNLRIIMIQYLLSGFELELYSMHYYIYIYWLSEFLYAW 540
Db 473 TMLLKQEPQORHCLAGTWLVYHNLRIIMIQYLLSGFELELYSMHYYIYIYWLSEFLYAW 532
QY 541 LMSTLSRADGSGMAEERIMEEQOGRSSKTKKKKKVRLPGLREITMSQAYQNMCAGMFKT 600
Db 533 LMSTLSRADGSGMAEERIMEEQOGRSSKTKKKKKVRLPGLREITMSQAYQNMCAGMFKT 592
QY 601 MVAEDMDGKVRKPKFELDSQVRYEHRFAPNSVMTPTPPVHYLPKEMSDLNKYSPPPQS 660
Db 593 MVAEDMDGKVRKPKFELDSQVRYEHRFAPNSVMTPTPPVHYLPKEMSDLNKYSPPPQS 652
QY 661 PELVYAAKSHFQQAQKMLENIPNDPHEVNRILKVAKNPVMVVKLLAGGHKKESKVPPEFD 720
Db 653 PELVYAAKSHFQQAQKMLENIPNDPHEVNRILKVAKNPVMVVKLLAGGHKKESKVPPEFD 712
QY 721 FSAHKYFPVVKLV 733
Db 713 FSVHKYFPVVKLV 725

RESULT 3
Q7T322
ID Q7T322 PRELIMINARY; PRT; 724 AA.
AC Q7T322
DT 01-OCT-2003 (TremBLrel. 25, Created)
DT 01-OCT-2003 (TremBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Sherman C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

```



```
RESULT 5
Q9W1A2 PRELIMINARY; PRT; 784 AA.
AC Q9W1A2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE CG4065 protein (GMO8606P).
GN CG4065.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Megaloptera; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Berkley;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Bencos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Douc L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W.,
RA Posler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maiti B., McIntosh T.C., McLeod M.F., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
[2]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuccio J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003463; RA047171.1; --
DR EMBL; AY060948; AAL28496.1; --
DR FlyBase; FBGN0034982; CG4065.
DR InterPro; IPR007244; Mak10.
DR Pfam; PF04112; Mak10; 1.
SQ SEQUENCE 784 AA; 89204 MW; DBDB966F0F523F17 CRC64;
Query Match 34.5%; Score 1326; DB 5; Length 784;
```

```
RA Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK026296; BAB15435.1; -.
KW Hypothetical protein.
SQ
SEQUENCE 211 AA; 24786 MW; F6E5BEFB6A6B77E CRC64;

Query Match
Best Local Similarity 29.0%; Score 1116; DB 4; Length 211;
Matches 211; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 523 MHEYYIYIYLSFELAYLWLMSTLSRADGQMAERIMEEQKGRSSKTKKKKVRPLSR 582
Db 1 MHEYYIYIYLSFELAYLWLMSTLSRADGQMAERIMEEQKGRSSKTKKKKVRPLSR 60
QY 583 EITMSQAYQNCAGMFKTMVAFDMGDKVRKPKFELDSQVYRHRFAPFNSVMTPPVHY 642
Db 61 EITMSQAYQNCAGMFKTMVAFDMGDKVRKPKFELDSQVYRHRFAPFNSVMTPPVHY 120
QY 643 LQFKEMSDLNKYSPPQSPPELYVAASHFQAKMTLENI PNPDEHVNRLKVAKPNFVVM 702
Db 121 LQFKEMSDLNKYSPPQSPPELYVAASHFQAKMTLENI PNPDEHVNRLKVAKPNFVVM 180
QY 703 KLLAGGHKESKVPPEFDFSAHKYFPVVKLV 733
Db 181 KLLAGGHKESKVPPEFDFSAHKYFPVVKLV 211

RESULT 7
Q8T3X5
ID Q8T3X5 PRELIMINARY; PRT; 708 AA.
AC Q8T3X5;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE AT26759p.
GN CG4065.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise B.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungali C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY089452; AAL90190.1; -.
DR FlyBase; FBgn0034982; CG4065.
DR InterPro; IPR007244; Mak10.
DR Pfam; PF04112; Mak10; 1.
SQ SEQUENCE 708 AA; 79791 MW; 181B7154C9C0C137 CRC64;

Query Match
Best Local Similarity 26.4%; Score 1017.5; DB 5; Length 708;
Matches 233; Conservative 104; Mismatches 214; Indels 67; Gaps 12;

QY 29 VYDITQDFEEACREKLGELHDKLFGFLPEASALEMMDPKMDAGVGNQVNRKVLNFEQ 88
Db 79 VYDITQDFEEACREKLGELHDKLFGFLPEASALEMMDPKMDAGVGNQVNRKVLNFEQ 138
QY 89 AIKDGTIKIKDLTLPBILGIMDTFCCLITWLEHSLAQVFTCLYIHNPDFIEDPAMKA 148
Db 139 ATATGAIKLDDLTSPSLIGIYDALFSLVSLWLEGNMDQVLTCLYLHAPAKIKOKALRV 198
QY 149 FALGILKICDIAREKVNKAARVEEDFQSMYTG---FKWANSVTDLRVTGMLKQVDDNQ 205
Db 199 FCTAVRNLIWIKKIITAAVAANVEEDFQ---LYGNSALLAAEAKAQATVYTSLLKQVDELI 256

QY 206 RRVKSTRSQGEERDPEVELEHQOCLAVGRVKFTRVLTLVLIATFKKTSAAVAEQ--- 262
Db 257 RKCKKLTST-----EDMVAVHRLRPMRHLFQVIYHVEQMASNDTVDDKVDI 303
QY 263 -KLMVQAADLLSALHSLHGIQAQNDTTKG-DHPIMMGPEPLVNQRLLPPTFPYAKII 320
Db 304 YKILLVASEMLPGIRNTLDRTQPE---KGSADAPNMGFSPIHDSRQPPAFPSIKIR 359
QY 321 KREEMVNYFARLDRIKTVCEVNLTLNLCILDFECFESBQS-PCVLSRSLQOTELVDN 379
Db 360 DRPSYQFLEEMISRFYACKTKYKYYGALNFIEYSKKSGQCILSRSLQILFSANM 419
QY 380 KKVFETHLQDMQDALRSFVDPVLPSPKCYLYNNHQAQKCIDSFVTHCV--RPFCSLIQ 437
Db 420 RMAHGKLPKQFLRHSQVFNSSPVLNAKHPVAADPKVQOQHLENFYCYINMTFTQIFR 479
QY 438 IHGHNARQDKLGHILEEFATLQDEMTFVFNRAEKVDAALHTMLLKQ-----E 487
Db 480 ICGFNARQDKLARLIENFDTIQ-----VDAARLDSMMNQLANERAMEGNE 526
QY 488 PQR---OHLACIGTWLYHNLRLIMIQVILSGFELELYSMHEYYIYIYWLSEFLYAWLMST 544
Db 527 PMATALKSHSTHFTSWLYNCPFRMLIFLMSGFELLYAVHEFLYIYWPYELIGFLVSA 586
QY 545 LSRADGQMAERIMEEQKGRS-----SKTKKKKKVRPLSRREITMSQAYQNM 593
Db 587 LTRTENILLAQEEYAEHQSKTQSGSGGAANKRKAAPKKNKKTQRPYRAEIVFYHALLSL 646
QY 594 CAGMFKTMVAFDMGDKVR 611
Db 647 CGMVKAMGALTQDGRV 664

RESULT 8
Q8BYJ9
ID Q8BYJ9 PRELIMINARY; PRT; 150 AA.
AC Q8BYJ9;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Corneal wound HEALING related protein homolog (fragment).
GN A330021G12RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Spinal cord;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK039316; BAC30316.1; -.
DR MGD; MGI:2441791; A330021G12RIK.
FT NON_TER 1
SQ SEQUENCE 150 AA; 17236 MW; 703BD7A706353982 CRC64;

Query Match
Best Local Similarity 20.1%; Score 772; DB 11; Length 150;
Matches 146; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 584 ITMSQAYQNCAGMFKTMVAFDMGDKVRKPKFELDSQVYRHRFAPFNSVMTPPVHYL 643
Db 1 ITMSQAYQNCAGMFKTMVAFDMGDKVRKPKFELDSQVYRHRFAPFNSVMTPPVHYL 60
QY 644 OFEKMSDLNKYSPPQSPPELYVAASHFQAKMTLENI PNPDEHVNRLKVAKPNFVVMK 703
Db 61 OFEKMSDLNKYSPPQSPPELYVAASHFQAKMTLENI PNPDEHVNRLKVAKPNFVVMK 120
QY 704 LLAGGHKESKVPPEFDFSAHKYFPVVKLV 733
```


Best Local Similarity 24.9%; Pred. No. 5.1e-38;
Matches 204; Conservative 134; Mismatches 302; Indels 179; Gaps 27;

QY 18 MPKXKSNNTNWVDITQDFEACRELKIGELLHDKLFGLEFAMSAIEMMDPKMDAGMIGN 77
DB 1 MFGAMESNES-EDVSKTFKCLDGLRGLGVTTTEHFRSLSDVMSAIELGEPKMDVG-VG- 57

QY 78 QVNRKVLNFEQAIKDGITIKIDLTLPILGIMTDFCCLITWLEGHSLAQVTFCLYHN 137
DB 58 --SKHITKLEAISGG-LYADD--YFQALAMDSLTAMVAVWLEGSALGSTVWTNVLUSN 112

QY 138 PDFIEDPAMKAFALGILKICIDIAREKVNKAAVFEE--EDFQ-SMTYGFK----- 183
DB 113 VTFVKHPVFHPFASGVNLFIRNAHALINSVGNLELPEDFNPQMLFSQRWAPRRVVQL 172

QY 194 MANSVTDLRVTGMLKDVEDDMQRRVKSTRSQGEERDEVELEHQOCLAVSRVKFTRVL 243
DB 173 MREQVTLTGTTG-----RKWRESAFSKQA-----YDICCavasRLEMFIML 213

QY 244 LTVL-----IAFTKKE 254
DB 214 LEITGLLVAPEIEDNFDHKVHLGDFYVHEXEYEGAEENSEKSEESTSDPTPSSEAKSE 273

QY 255 TSAAVAQKLMVQAD-----LLSAIHNSLHHGIIQAO 287
DB 274 TTEKKE-QKSNDKATDETDQNDENVGDEDEDDMKQKPNFGFASILAERLCKVTKAYA 332

QY 288 DTTK-----GHPIMMGPELVNQRLLPPTPRYAKITIKREEMVNYFARLIDR 335
DB 333 ETVGLGRAPDNIDGDSYLVLGAYEPKTCIRNIPACFPKIKVPSROEAADWWVKCAER 392

QY 336 IKTVCEVNLNLHC-ILDFCFEFSQPCVLSRLQ-TTFLVNKKVFGTHL----- 387
DB 393 IYHLCIVTPKYSKDLNLYFYFARTFGQACVFRSLQLICMPVDN-----HLCGENR 446

QY 388 -MODVMKDALSFDVPPVLSKCYLYNNHQAQKDCDSFVTHCVRPFCSLIIQIHGHNARQ 446
DB 447 SIADAVEYSLNCFQPIILDRVSPYKQTAGSLVFLFNHMSKLAIIVYSGFGCNLSRQ 506

QY 447 RDKLGHILEEPAATLODEFTVFYFNRAEK--VDAALHTMLLKQEPORQHLACLTGTVLHN 504
DB 507 RDRLEWATEDLGOIHS-----YAGLEERTDEVLSSGMVTAKEQNSYHVSATVFVHNL 561

QY 505 LRIMTOYLLSGPELYLSMHEYYTYWYLSFLYALWMLSTLSRADGSQMAERIMEEQOK 564
DB 562 LAIINHIFELGRMDLYVDPYEPYIYWFISGVOAHMRTILERS-----QEIQLNVYQ 614

QY 565 GRSSKTKKKK-----KV-RPLSREITMSQ-----AYQNCAGMEKMTVAFDMGKV 610
DB 615 ANPLRETKNKLWESECKLGBELKGRVAHQFVLNQIAISMLSDGVVRLTVVLIRKGII 674

QY 611 RKPFPKELDEQVRYHRFAPFNSVMTPPVHYLOPKEMSDLNK-YSPPPQSPELYVAASK 669
DB 675 KMPKGGDDAKLRFRRFPFSLGPPVVDYVERFKSDSGIDQMYBDKJET--LIDQAOK 732

QY 670 HFQQAQKMLENIPN---PDHEVNRILKVAKPNFVVMKLL 705
DB 733 SPNEAREHLEKIDNSVEQREMQLIHVAKNSIIACRVL 771

RESULT 11
Q8WZ5 PRELIMINARY; PRT; 908 AA.
ID Q8WZ5
AC Q8WZ5
DT 01-MAR-2002 (TEMBLrel. 20, Created)
DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
DE Conserved hypothetical protein.
GN B7N14.110.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
NCBI_TaxID=5141;

[1]
SEQUENCE FROM N.A.
RP Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL669986; CAD21048.1; --
DR InterPro; IPR007244; Mak10.
KW Pfam; PF04112; Mak10; 1.
KW Hypothetical protein.
SQ SEQUENCE 908 AA; 101120 MW; D40AF51D21542FF6 CRC64;

Query Match 9.8%; Score 377.5; DB 3; Length 908;
Best Local Similarity 21.5%; Pred. No. 1.4e-22;
Matches 162; Conservative 106; Mismatches 277; Indels 209; Gaps 24;

QY 31 DITODFEACRELKIGELLHDKLFGLEFAMSAIEMMDPKMDAGMIG-----NQVNRK 82
DB 10 NITKFLSACKLTETGEIVKDGFTLFEAVGALEIGDKMDSGCLAPGETLEETVDVNRQ 69

QY 83 VLNFEQAIKDGITIKIDLTLPILGIMTDFCCLITWLEGHSLAQVTFCLYI----- 135
DB 70 -----LSAPEVLGIIDQLLCLEMAWHLGYPLAQITLTSVYIEALVEPA 112

QY 136 -----HNPDFI-----EDP---AMKAFALGILKICIDIAREKVNKAAVFEEDFQSMYGF 182
DB 113 PATLOEADFVRNKRTPRDEMTVLRAYCIGLVRTCADVLEITRDELYEEDFVTNTRR 172

QY 183 KMANSVTDLRVTGMLKDVEDDMQRRVKSTRSQGEERDEVELEHQOCLAVFSRVKTRV 242
DB 173 NLMHIDRYEI-----RDEILSAKHTHELRAEIGTD----- 204

QY 243 LITVLIAFTKETSAAVAQKLMVQ--AADLLSAIHNSLHHGIIQAOQNTTKGDHPIMGF 300
DB 205 -ITQALSFRLELSAFLRALELIEDRGSPDSLQIPWYQMSVWEAILKPLGLAKEVPEAF 263

QY 301 EPLVNQRLLPPTPRYAKITIKREEMVNYFARL-----IDRIKTVCEVNLNLHCILDFPC 356
DB 264 STKLQKCLASSPPRPVITLSFDEAAKHFKLCSAIDAIAK-----ILDVHDSQSLNFFV 319

QY 357 EFSEQSP-----CVLSKSLIQTTFVNDKKVFGTHLMQDMVKDALRSFVDP--PVLSP 407
DB 320 HEQAQKPPQVYIRCLLQNLFLKQNVLLDLSI-----RQWDDDLSTAVMPAHELLOP 373

QY 408 KCYLNN-----HOAKDCI-----DSFVTHCVRPFCSLIIQIHGHNARORDKLGHILEP 457
DB 374 -----NNDLVEAPHSAFAIAHQMBELFRKRAQAVIDIPVLQCNRCRVRRLCHLIQDW 428

QY 458 ATLQDEFMTFYENRAEKVDAALH-----TMLLQEPQROHLACLTGTVLVLH 503
DB 429 EQVOLD-----AEDIDQLLQIODEKPLAYQSQTTLTGSEP--GVSLEPSSWAYLY 478

QY 504 NLRIMIQVLLSGFELYLSMHEYYTYWYLS-----EFLYAWLMSLTS--RADGS 551
DB 479 KLRIMEWIVQLGFLELYVAPDELAGWYWLSHLAKTRAQHVERIQAFTHRFSLSRSSS 538

QY 552 QMAERIMEEQOKGRSSKTKKKKKVRLPSRITMSQAYQ-----N 592
DB 539 SSTSTSPSPSSSNPSPSSYTKPPIPATPWIPPTQPVHPRGWPFVSISSSSLSGPSFT 598

QY 593 MCAGMFKTM-----VAFDMD-----GKVRK 612
DB 599 PCAGTGTGTTTTHYSTTPEAAFTTSYLRSTILDRAITWEFADTVSVIYTLISRLGLITP 658

QY 613 PKFELDEQVRYHRFAPFNSVMTPPVHYLOPK 646
DB 659 PRPYGTDEKRYEVRMKPFRVSLPEVPSYTTWK 692

RESULT 12


```
DR EMBL; AL929356; CAD51797.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1218 AA; 147226 MW; 7A081A9B4D0E10ED CRC64;

Query Match      3.6%; Score 150; DB 5; Length 1218;
Best Local Similarity 19.4%; Pred. No. 0.0027;
Matches 131; Conservative 106; Mismatches 216; Indels 224; Gaps 36;

Qy 42 ELKUGELHDKLFGLEFAMSATMDPKM-----IGNQNRKVLNFEQAIDGT 94
Db 22 DLKDEEIKQDFSVKRVKVCSEGLGDKLDCGIIHPEKRLSIIDCEERYINDE----- 74
Qy 95 IKIKDLTLPGLIGMDTCFCCLITWLEHSLAQVFTCLYH-----NPDFIEDPA 145
Db 75 -----LSYDDIIFIIDHLLMLKLGKGNHPLFTVLSYCYLHDSKVINCCKNGFLBEVF 129
Qy 146 MK--AFALGILKICDIAREKVNAAV--FBEEDFQSMYTGFKMANSYTDLRVTGMLKDV 200
Db 130 MKWLAAFSYDKIYM-DIFKND-NKEFIKYIEKHDDQKKNFN--KNKVADNK-----KDS 179
Qy 201 EDD--MORRVKSTRSQGEERDPVELEHQOCLAVFSRVKTRVLLTVLIAFTKETSAV 258
Db 180 NQDNNMEXKKE-----ESEEKEIEE-----KKSDTI 207
Qy 259 ABAQKLMVQAADLLSAIHSNHHGIIQAQNDTTKGDHPIMMGFEPLVNQRLPPTPRYAK 318
Db 208 NIKYK-----NNIHDNICENTYKN-----ICEEDYVENKMP--LEFQLFL 248
Qy 319 II--KREEMVNYFA--RLIDRIKTVCBVNL-----NL-----HCILD 353
Db 249 IFYSSSELIDYVIQNNFVHRDDYKCGLVNLDCLLYHCNRRAVILKNLFIKRCFTK 308
Qy 354 FCFESFESQPCVLSRLIQTFLVNKK-----VEGTHLMQDMVVDALRS 398
Db 309 PFTYEE-----NKKNKDIFYILKRIKFIITYFTSLNQLIPDMCES 350
Qy 399 FVDPVPLSPKCVLYNHQAQKIDCSFVTHCVRPFCGLIQIHGHNARORDKLGHILEFA 458
Db 351 NKDS--IKENC-----KQILNCINIDNDM--NCNI-----NEKGVNKNKFMQIEE 395
Qy 459 TLQDE---FMTPYFNAEKVDAAALHTMLLKQBPQCHLACLTG---WVLYHNLIRIMIQL 512
Db 396 BEEKEKKKKKLYFNK-----YFLMYKTHVSKHVTKLTSISEGYLFYKNIITDIQVI 447
Qy 513 LSGFLELYSMHE-----YYIYIY-----LSEFLYAWLMTLSRADSGMA---E 555
Db 448 AD--YIKLTNMKSPFDIKNHLHYLFKYSKSNILIKCIFKCLTQIIQVEKKEFAKPEK 505
Qy 556 BRIMEEQOQGRSSKTKKKKVRPLSREITMSQAYQNM-----AGMKVMVAFDMQKVRK 612
Db 506 EYLITQOSK--ASKHLNNKK-----NVESGKNNSCPIIANKENTKNEKREDLES DK 555
Qy 613 PK-----FELDS 620
Db 556 QKQNIYDDVEIFESDDE 572

RESULT 15
Q8JGR7 PRELIMINARY; PRT; 964 AA.
AC Q8JGR7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2003 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DKFZP3434B168-like protein.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22035902; PubMed=12006978;
```

RA Golling G., Amsterdam A., Sun Z., Antonelli M., Maldonado E., Chen W.,
RA Burgess S., Haldi M., Artzt K., Farrington S., Lin S.-Y., Nissen R.M.,
RA Hopkins N.;
RT "Insertional mutagenesis in zebrafish rapidly identifies genes
RT essential for early vertebrate development.";
RL Nat. Genet. 31:135-140(2002).
RL EMBL; AY099529; AAM28217.1; -.
DR InterPro; IPR00938; ARM.
SQ SEQUENCE 964 AA; 106914 MW; 280310F72EADAAA CRC64;

Query Match 3.6%; Score 139.5; DB 13; Length 964;
Best Local Similarity 19.0%; Pred. No. 0.015;
Matches 182; Conservative 123; Mismatches 296; Indels 355; Gaps 50;

Qy 35 DFEACRELKGE-----LHDKLFGLEFAMSATMDPKM-DAGMIGNQNRKVLNFEQ 88
Db 28 ELDKGLSKLGECEAVVLPFKLPQKYPFFILINSAPFLKLAIFRLGN-----NF-- 78
Qy 89 AIKDGTTIKDLTLPGLIGMDTCFCCLITWLEHSLAQVFTCLYHNPDIEDPAMKA 148
Db 79 -LRCLVKVLTQSEKHLKILNV-----DEFVKRVFS-VIHS---NDPVARA 120
Qy 149 FALGIL-KICDIAEKVN-----KAAVEEEDPQSMYTGFK--MANSV 188
Db 121 IITRLMGLSLASIIIPERKNAHHSIRQSLDSDHNVVEARAIFAAASFSSHSDFAAGICNKI 180
Qy 189 TDLRVTGMLKQVEDDMQRRVKSTRSQGEERDPVELEHQOCLAVFSR-----VKF 239
Db 181 SE-----MIQGLDTFVELKGLIPLMQ-----HMHDSLASCSRELQELVSSVPS 227
Qy 240 TRVLITVLIAITKETSAAVE-----AQKLMVQAADLLS----- 273
Db 228 TSMILVLTHTTQLATSLVDIPEQICLLQYLKEDPRKAVKRLSIQDLKLLAKKAPHLW 287
Qy 274 -----AIH---NSLHHGIIQAQNDTTKGDHPIMMGFEPLVNQRLPPTPRYAKI 319
Db 288 TRKNIQVLCCEALHTPPYNSLGLMGLSVLSTGTTAIKQYFSP--NAGDSSPA-PHHTDL 344
Qy 320 IKREEMVNYFAEL-----IDRIKTVG---EVVNLTNLHCLIDFFCFESEQSPCVLS 367
Db 345 VKLAQECCHSDSLAVAAGHITVLTSLIAAFCEPEKVIQI-----EQTTVMGM 390
Qy 368 RSLIQTTFELVNKKVFGTHLMQDMVKDALRSFDPVPLSPKCVLYNHQAQKIDCSFV-- 425
Db 391 ESLIILCSQDSDSKTA-----QATLKALTSLVQMLKTCF-----HLSQSSVELLRQ 437
Qy 426 THCVRPFC-----SLIQIHGHNARORDKLGHILEF-----ATLQDEM--- 465
Db 438 LHCA---CDPARVLMQALAAIATQPPVLVEGMLGDLLELFRVASHRTSEKQOELLVSLA 494
Qy 466 TFYFNAEKVDAAALHTMLLKQBPQCHLACLTGTVLYHNLIRIMIQVLLSGPELELYS-- 522
Db 495 TVLVFASQASAEVKAIVRQ--QLENVA--NGWTVIQIARQASRMGCHDFSRILEYQSLR 550
Qy 523 ---MHEYYIYIYWY--LSEFL-----YAWLMTLSRADSGSQMAERIMEEQOK 564
Db 551 TRVASEHYF-FWLNSLMEFSQAECLSGLEDGYSAAMSASEA-----LKSQYK 599
Qy 565 GRSSKTKKKKKVRPLS-----REITMSQAYQNM----- 594
Db 600 GIAS-LTAASTPLSPLTFOCFEVKRLIDTLQALSQILCTCNLSLKTSPPPAIATTIALSSG 658
Qy 595 -----AGMKTMVAFDM-----GKVRKPELDSQVR----- 623
Db 659 SLDQRCGRISTQMKFSMDFRSLAARYADLYOSSDADYATLURNVELQQOCLLYSVYLE 718
Qy 624 -----YEHR-FAPFNSVMTP-----PPVHYLQ----- 644
Db 719 ALLIDPQTASFQFEGTHGSILAESEYELRMVAFVNHVLEVEVENSRKHHPVSYLHTGCLC 778
Qy 645 -----FKMESDLN---KYSPPPOSPELYVAASKHFQOAKMILENI--- 681
Db 779 DTVAIIKIPLSFQRYFFQKLOSTSIKLALSPSPRTNPEPIPVQNN-QQLTLKVEGVIQH 837

Qy 682 ---PNPDHEVNRI-LKVA-----KP--NFVVMKLLAGGHKKESKVPPEFDPSAHKY 726
Db 838 GSSPGLFRKIQAVCLKVSSTLQTKPGSDFKI-PLESKTNEIEQKVEPHNDYESTQF 892

Search completed: August 10, 2004, 21:51:28
Job time : 47 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 21:44:16 ; Search time 58 Seconds
(without alignments)

3570.816 Million cell updates/sec

Title: US-10-001-857-145

Perfect score: 3849

Sequence: 1 MWKASVDDDDSGWELSMPE.....KVPPEPDSAKHYFPVVKLV 733

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq_29Jan04.*

1: geneseqp1980s.*

2: geneseqp1990s.*

3: geneseqp2000s.*

4: geneseqp2001s.*

5: geneseqp2002s.*

6: geneseqp2003as.*

7: geneseqp2003bs.*

8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3849	100.0	733	5	ABP52902 Human lun
2	3783	98.3	725	5	AAO19400 Human mol
3	3728.5	96.9	726	7	ADC31012 Human nov
4	1326	34.5	784	4	ABE60388 Drosophil
5	1100	28.6	239	5	ABG69816 Human sec
6	685	17.8	160	7	ADC32736 Human nov
7	127	3.3	733	3	AAO07449 Amino aci
8	127	3.3	733	3	AAO07448 Amino aci
9	127	3.3	761	6	AAE36464 Human col
10	127	3.3	761	7	ADD14191 Human src
11	127	3.3	1089	2	AAW01896 Nonsense-
12	127	3.3	1089	2	AAO5835 Yeast Nmd
13	127	3.3	1089	3	AAO58055 Yeast Nmd
14	126	3.3	1257	4	ABG17335 Novel hum
15	121.5	3.2	1184	4	ABG02338 Novel hum
16	120	3.1	644	4	AAE95305 Human pro
17	119	3.1	1488	7	ADC31044 Human nov
18	119	3.1	3353	5	AAU82706 Amino aci
19	119	3.1	3412	7	ADBE6580 Human ubi
20	118.5	3.1	5303	4	ABE67866 Drosophil
21	116	3.0	736	5	ABE65160 Hypoxia-r
22	115.5	3.0	1141	4	ABG17210 Novel hum
23	115	3.0	633	5	ABG96286 Human ova
24	115	3.0	912	4	AAO90771 Human she
25	115	3.0	912	5	ABG34124 Human non

ALIGNMENTS

RESULT 1

ABP52902
ID ABP52902 standard; protein; 733 AA.

XX AC ABP52902;

DT 05-NOV-2002 (first entry)

DE Human lung specific protein sequence SEQ ID NO:145.

XX Human; lung; lung specific nucleic acid; L5NA; lung specific protein;

KW LSP; cytostatic; gene therapy; vaccine; metastasis; lung cancer;

KW squamous cell carcinoma.

XX Homo sapiens.

XX WO200264788-A2.

XX 22-AUG-2002.

XX 20-NOV-2001; 2001WO-US045080.

XX 20-NOV-2000; 2000US-0252054P.

XX (DIAD-) DIADEXUS INC.

XX Macina RA, Recipon H, Chen S, Sun Y, Liu C;

XX WPI; 2002-657601/70.

XX New lung specific nucleic acid useful in gene therapy or as vaccines for treating lung cancer (e.g. squamous cell carcinoma) or non-carcinoma lung diseases, as well as for diagnosing, monitoring or staging these diseases.

XX Claim 11; Page 239-242; 282pp; English.

XX The present invention describes an isolated lung specific nucleic acid (LSNA) comprising a sequence that: (a) encodes any of the 93 amino acid sequences comprising 17 - 733 amino acids, given in ABP52873 to ABP52965; (b) comprises any of 115 sequences comprising 148 - 3193 base pairs (bp); given in ABQ75262 to ABQ75376; (c) selectively hybridises to (a) or (b); or (d) has 60 % sequence identity to (a) or (b). LSNA and lung specific protein (LSP) sequences have cytostatic activity and can be used in gene therapy and vaccines. LSNA and LSPs are useful for diagnosing and monitoring the presence and metastases of lung cancer in a patient. An antibody that specifically binds to an LSP can be used for determining the presence of an LSP in a sample, as well as for treating a patient

Abu89754 Protein d
Aab63240 Human bre
Abp52168 Mouse pot
Add37507 Mouse pho
Add37469 Mouse pho
Abg06288 Novel hum
Abg15642 Novel hum
Abg23670 Novel hum
Abg06619 Novel hum
Abg05642 Novel hum
Abj19343 NOVA rela
Abg14789 Novel hum
Abg06938 Novel hum
Abg02332 Novel hum
Abg10803 Novel hum
Abg06060 Novel hum
Abg06606 Novel hum
Abg09642 Novel hum
Abg07064 Novel hum
Aab18324 Plasmodiu

115 3.0 912 6 ABU89754
115 3.0 937 4 AAB63240
115 3.0 1187 5 ABP52168
115 3.0 1187 7 ADD37507
115 3.0 1187 7 ADD37469
115 3.0 1565 4 ABG06288
114.5 3.0 1029 4 ABG15642
114.5 3.0 1190 4 ABG23670
114.5 3.0 1190 4 ABG06619
114.5 3.0 1190 4 ABG05642
114.5 3.0 1191 6 ABJ19343
114 3.0 1304 4 ABG14789
114 3.0 985 4 ABG06938
114 3.0 1065 4 ABG02332
114 3.0 1078 4 ABG10803
114 3.0 1078 4 ABG06060
114 3.0 1078 4 ABG06606
114 3.0 1078 4 ABG09642
114 3.0 1261 4 ABG07064
114 3.0 1558 3 AAB18324

CC with lung cancer, particularly by inducing an immune response against the
 CC lung cancer cell expressing the LSNAs or LSPs. In particular, these LSNAs
 CC and LSPs are useful for identifying, diagnosing, monitoring, staging, and
 CC imaging and treating lung cancer (e.g. squamous cell carcinoma) and non-
 CC cancerous disease states in lung
 XX
 SQ Sequence 733 AA;

Query Match 100.0%; Score 3849; DB 5; Length 733;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 733; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVMKASVDDDDSGWELSMPEKMEKSNNTNWDTITQDFEACRELKLGELLHDKLFLGLFEAM 60
 DB 1 MVMKASVDDDDSGWELSMPEKMEKSNNTNWDTITQDFEACRELKLGELLHDKLFLGLFEAM 60

QY 61 SAIEEMDPKMDAGMIGNQVNRKVLNFEQAIDGTIKIKDITLPELIGIMDTFCCLITWL 120
 DB 61 SAIEEMDPKMDAGMIGNQVNRKVLNFEQAIDGTIKIKDITLPELIGIMDTFCCLITWL 120

QY 121 EGHSLAQTFTCLYIHNPFDIEDPAMKAFALGILKICDIAREKVNKAAVFEEDFQSMY 180
 DB 121 EGHSLAQTFTCLYIHNPFDIEDPAMKAFALGILKICDIAREKVNKAAVFEEDFQSMY 180

QY 181 GFKMANSVTDLRVTKMLKDVDDMORRVKSTRSQGERDPEVELEHQOCLAVFSRVKFT 240
 DB 181 GFKMANSVTDLRVTKMLKDVDDMORRVKSTRSQGERDPEVELEHQOCLAVFSRVKFT 240

QY 241 RVLTLVLAFTKETSVAEAKLMQVQADLLSAHNSLHGICQANDTTKGDHPIMMGF 300
 DB 241 RVLTLVLAFTKETSVAEAKLMQVQADLLSAHNSLHGICQANDTTKGDHPIMMGF 300

QY 301 EPLVNQRLPPTFPFYAKIIKREEMVNYFARLIDRIKTVCEVWNLNLHCLIDFCEPSE 360
 DB 301 EPLVNQRLPPTFPFYAKIIKREEMVNYFARLIDRIKTVCEVWNLNLHCLIDFCEPSE 360

QY 361 QSPCVLSRLQTTFLVNDKVKVFGTHLMQDMVKQALRSFVDPVLSPKCYLYNNHQAKDC 420
 DB 361 QSPCVLSRLQTTFLVNDKVKVFGTHLMQDMVKQALRSFVDPVLSPKCYLYNNHQAKDC 420

QY 421 IDSFVTHCVRFCSLIQIHGHNARQDKLGHILEEPATLODEFWTFYFNRAEKVDAALH 480
 DB 421 IDSFVTHCVRFCSLIQIHGHNARQDKLGHILEEPATLODEFWTFYFNRAEKVDAALH 480

QY 481 TMLLKQEPQOHLACLGTVLVLNLRIMIQYLLSGFELELYSMHEYYIYVWLYSEFLYAW 540
 DB 481 TMLLKQEPQOHLACLGTVLVLNLRIMIQYLLSGFELELYSMHEYYIYVWLYSEFLYAW 540

QY 541 LMSTILSRADGSGMAEERIMEEQKGRSKTKKKKKVRPLSREITMSQAYQNMCAGMFKT 600
 DB 541 LMSTILSRADGSGMAEERIMEEQKGRSKTKKKKKVRPLSREITMSQAYQNMCAGMFKT 600

QY 601 MVAFDMDGKVRKPKFELDSQVRYEHRFAPNSVMTPPVHYLOFKEMSDLNKYSPPQS 660
 DB 601 MVAFDMDGKVRKPKFELDSQVRYEHRFAPNSVMTPPVHYLOFKEMSDLNKYSPPQS 660

QY 661 PELVAAKSHFQQAQMIENIPNDPHEVNRILKVAKENFVVMKLLAGGHKESKVPPEFD 720
 DB 661 PELVAAKSHFQQAQMIENIPNDPHEVNRILKVAKENFVVMKLLAGGHKESKVPPEFD 720

QY 721 FSAHKYFPVVKLV 733
 DB 721 FSAHKYFPVVKLV 733

RESULT 2

AA019400

ID AA019400 standard; protein; 725 AA.

XX AC AA019400;

XX DT 10-DEC-2002 (first entry)

XX

XX

DE
XX
KW
KW
KW
XX
OS
XX
PN
XX
PD
XX
PF
XX
PR
PR
PR
PR
PR
PR
PR
PR
XX
PA
XX
PI
PI
PI
PI
XX
DR
DR
XX
PT
PT
PT
PT
XX
PS
XX
CC
CC
CC
CC
CC
CC
CC
CC
CC
XX
SQ

Human molecule for disease detection and treatment protein #3.
 Human; molecule for disease detection and treatment; MDDT; gene therapy;
 cytostatic; antiarteriosclerotic; hepatotropic; anti-HIV; antiallergic;
 antiinflammatory; antiasthmatic; cerebroprotective; nootropic;
 neuroprotective; antiparkinsonian; cardiant; antianginal.

Homo sapiens.

WO200270709-A2.

12-SEP-2002.

08-FEB-2002; 2002WO-US003709.

09-FEB-2001; 2001US-0268117P.

15-FEB-2001; 2001US-0269618P.

23-FEB-2001; 2001US-0271118P.

07-MAR-2001; 2001US-0274486P.

09-MAR-2001; 2001US-0274436P.

28-NOV-2001; 2001US-0334229P.

01-FEB-2002; 2002US-0353284P.

(INCY-) INCYTE GENOMICS INC.

Lal PG, Baughn MR, Yao MG, Wallia NK, Elliot VS, Xu Y;

Honchell CD, Yue H, Ding L, Gietzen KJ, Ison CH, Lu DAM;

Hafalia AJA, Ghandi AR, Thangavelu K, Sanjanwala MM, Tang YT;

Runkumar J, Griffin JA, Swarnaker A, Azimzai Y, Sapperstein SK;

Burford N, Lee BA, Lu Y, Tran UK, Marquis JP;

WPI; 2002-713453/77.

N-PSDB; AAL49929.

New human molecules for disease detection and treatment (MDDT), useful

for diagnosing, treating and preventing diseases or conditions associated

with the aberrant MDDT expression, e.g. cancer, AIDS, asthma, diabetes,

hepatitis.

Claim 1; Page 128-130; 177pp; English.

The present invention relates to human proteins and coding sequences of

molecules for disease detection and treatment MDDT. The sequences can be

used in the treatment of diseases associated with the decreased

expression or overexpression of MDDT, such as cell proliferative (cancer,

atherosclerosis, hepatitis), autoimmune/inflammatory (e.g. AIDS, renal

allergies, Addison's disease, asthma), developmental (dwarfism, renal

tubular acidosis), neurological (e.g. stroke, Parkinson's disease,

epilepsy) and cardiovascular (congestive heart failure, myocardial

infarction, angina pectoris) disorders. The present sequence is a protein

of the invention

Sequence 725 AA;

Query Match 98.3%; Score 3783; DB 5; Length 725;

Best Local Similarity 98.8%; Pred. No. 0;

Matches 724; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

QY 1 MVMKASVDDDDSGWELSMPEKMEKSNNTNWDTITQDFEACRELKLGELLHDKLFLGLFEAM 60

DB 1 MVMKASVDDDDSGWELSMPEKMEKSNNTNWDTITQDFEACRELKLGELLHDKLFLGLFEAM 60

QY 61 SAIEEMDPKMDAGMIGNQVNRKVLNFEQAIDGTIKIKDITLPELIGIMDTFCCLITWL 120

DB 61 SAIEEMDPKMDAGMIGNQVNRKVLNFEQAIDGTIKIKDITLPELIGIMDTFCCLITWL 120

QY 121 EGHSLAQTFTCLYIHNPFDIEDPAMKAFALGILKICDIAREKVNKAAVFEEDFQSMY 180

DB 121 EGHSLAQTFTCLYIHNPFDIEDPAMKAFALGILKICDIAREKVNKAAVFEEDFQSMY 180

QY 181 GFKMANSVTDLRVTKMLKDVDDMORRVKSTRSQGERDPEVELEHQOCLAVFSRVKFT 240

DB 181 GFKMANSVTDLRVTKMLKDVDDMORRVKSTRSQGERDPEVELEHQOCLAVFSRVKFT 240

QY 241 RVLTLVLIATFKKTSVAEAKLMVQAADLLSAIHNSLHHGIOAONDTTKGDHPIMMGF 300
DB 241 RVLTLVLIATFKKTSVAEAKLMVQAADLLSAIHNSLHHGIOAONDTTKGDHPIMMGF 300
QY 301 EPLVNQRLPPTFFRYAKIIKREEMVNYFARLIDRIKTVCEVNLNLHCHILDFCFE 360
DB 301 EPLVNQRLPPTFFRYAKIIKREEMVNYFARLIDRIKTVCEVNLNLHCHILDFCFE 360
QY 361 QSPCVLSRLSLQTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVLSPKCYLYNNHQAKDC 420
DB 361 QSPCVLSRLSLQTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVLSPKCYLYNNHQAKDC 420
QY 421 IDSFVTHCVRPFCSLIQIHGHNARQDKLGHILEEFATLODEFWTFYFNAEKVDALH 480
DB 421 IDSFVTHCVRPFCSLIQIHGHNARQDKLGHILEEFATLODE-----AEKVDAALH 472
QY 481 TMLLKQBPQORHLACLGTVWLYHNLRIIMQYLLSGFELELYSMHEYYIYVWYSEFLYAW 540
DB 473 TMLLKQBPQORHLACLGTVWLYHNLRIIMQYLLSGFELELYSMHEYYIYVWYSEFLYAW 532
QY 541 LMSTLSRADGSQMAEERIMEEQKGRSSKTKYKKKKVRLPSREITMSQAYQNMCGMFKT 600
DB 533 LMSTLSRADGSQMAEERIMEEQKGRSSKTKYKKKKVRLPSREITMSQAYQNMCGMFKT 592
QY 601 MVAEDMDGKVRKPELDSQVRYEHRFAPNSVMTPPVHVLOPKEMSLNKYSPPPQS 660
DB 593 MVAEDMDGKVRKPELDSQVRYEHRFAPNSVMTPPVHVLOPKEMSLNKYSPPPQS 652
QY 661 PELVVAASKHFQQAAMILENIPNDPHEVNRILKVAKPNFVVMKLLAGGHKESKVPPPEF 720
DB 653 PELVVAASKHFQQAAMILENIPNDPHEVNRILKVAKPNFVVMKLLAGGHKESKVPPPEF 712
QY 721 FSAHKYFPVVKLV 733
DB 713 FSAHKYFPVVKLV 725

RESULT 3
ADC31012
ID ADC31012 standard; protein; 726 AA.
XX AC ADC31012;
XX DT 18-DEC-2003 (first entry)
XX DE Human novel polypeptide sequence, SEQ ID NO:1094.
XX KW Human; diagnostic; drug screening; forensics; gene mapping;
KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;
KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
KW ulcers; osteoporosis; autoimmune disease; cancer;
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;
KW neuroprotective; anti-anaemic; anticoagulant; thrombolytic; vulnerary;
KW anti-ulcer; osteopathic; immunosuppressive; anti-inflammatory; cytostatic;
KW gene therapy; chromosome 9.
XX OS Homo sapiens.
XX PN WO2003029271-A2.
XX PD 10-APR-2003.
XX PF 24-SEP-2002; 2002WO-US030474.
XX PR 24-SEP-2001; 2001US-0324631P.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
PI Haley-Vicente D, Drmanac RT;
XX

DR WPI; 2003-371981/35.
DR N-PSDB; ADC30041.
XX New polynucleotide and polypeptide useful for diagnosing, preventing or
PT treating conditions such as neurodegenerative diseases, anemias, platelet
PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
PT cancer.
XX Claim 20; SEQ ID NO 1094; 1185pp; English.
PS The invention relates to 971 novel human cDNA sequences (ADC29919-
XX ADC30899) and the polypeptides they encode (ADC30890-ADC31860). The
CC invention also relates to nucleic acid sequences over 99% identical with
CC the novel human cDNAs. The invention additionally encompasses expression
CC vectors and host cells comprising a nucleic acid of the invention; the
CC recombinant production of a polypeptide of the invention; an antibody
CC against a polypeptide of the invention; a method of detecting
CC polynucleotides or polypeptides of the invention; and methods of
CC identifying a compound which binds to a polypeptide of the invention.
CC The invention further discloses methods of preventing, treating or
CC ameliorating a medical condition; kits comprising polynucleotide probes
CC and/or monoclonal antibodies for carrying out the methods of the
CC invention; methods for the identification of compounds that modulate the
CC expression or activity of the polynucleotide and/or polypeptide; and 767
CC contig sequences corresponding to the cDNA sequences of the invention
CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
CC -ADC33394). The nucleic acids and polypeptides of the invention are
CC useful in diagnostics, drug screening, forensics, gene mapping, in the
CC identification of mutations responsible for genetic disorders or other
CC traits, for assessing biodiversity, and in producing many other types of
CC data and products dependent on DNA and amino acid sequences. They are
CC also used for treating diseases such as Parkinson's disease, Alzheimer's
CC disease and other neurodegenerative diseases, anaemia, platelet
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
CC cancer. The nucleic acids may also be used as hybridisation probes or
CC primers, and in the recombinant production of a protein. The polypeptides
CC are also useful in generating antibodies, as molecular weight markers,
CC and as food supplements. The present sequence represents a specifically
CC claimed human polypeptide sequence of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 726 AA;
Query Match 96.9%; Score 3728.5; DB 7; Length 726;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 718; Conservative 1; Mismatches 6; Indels 9; Gaps 2;
QY 1 MVMKASVDDDDSGWELSMPEKMEKSNNTNWVDITQDFEACRELKGLGELHDKLFGLEFAM 60
DB 1 MVMKASVDDDDSGWELSMPEKMEKSNNTNWVDITQDFEACRELKGLGELHDKLFGLEFAM 60
QY 61 SAIEWMDPMQDAGMTGNQVRKVLNFEQAIDKGTIKIKDLTLPGLIGIMDTCCCLITWL 120
DB 61 SAIEWMDPMQDAGMTGNQVRKVLNFEQAIDKGTIKIKDLTLPGLIGIMDTCCCLITWL 120
QY 121 EGHSLAQVFTCLYTHNPDTEDPAMKAPALGILKICDIAREKVNKAAPFEEDFQSMY 180
DB 121 EGHSLAQVFTCLYTHNPDTEDPAMKAPALGILKICDIAREKVNKAAPFEEDFQSMY 180
QY 181 GFKMANSVTDLRVTGMLKDVEDDMORRVKSTRSRQGEERDPEVELEHQOCLAVFSRVKFT 240
DB 181 GFKMANSVTDLRVTGMLKDVEDDMORRVKSTRSRQGEERDPEVELEHQOCLAVFSRVKFT 240
QY 241 RVLTLVLIATFKKTSVAEAKLMVQAADLLSAIHNSLHHGIOAONDTTKGDHPIMMGF 300
DB 241 RVLTLVLIATFKKTSVAEAKLMVQAADLLSAIHNSLHHGIOAONDTTKGDHPIMMGF 300
QY 301 EPLVNQRLPPTFFRYAKIIKREEMVNYFARLIDRIKTVCEVNLNLHCHILDFCFE 360
DB 301 EPLVNQRLPPTFFRYAKIIKREEMVNYFARLIDRIKTVCEVNLNLHCHILDFCFE 360

QY 361 QSPCVLSRLSLLQTFLVNNKVFVTHLMQDMVKDALRSFVD-PPVLSPKCYLYNNHQAKD 419
 Db 361 QSPCVLSRLSLLQTFLVNNKVFVTHLMQDMVKDALRVFCQILRVLSPKCYLYNNHQAKD 420
 QY 420 CIDSVTHCVRPFCSLIQIHGHNRRARDKLGHIIEEFATLQDEFMTFFNRAEKVDAAL 479
 Db 421 CIDSVTHCVRPFCSLIQIHGHNRRARDKLGHIIEEFATLQDE-----AEKVDAAL 472
 QY 480 HTMLLKQBPQRQHLACLGTTWLYHNLRIIMIOYLLSGFELELYSMHEYIYIYWLSEFLYA 539
 Db 473 HTMLLKQBPQRQHLAWLGTWLYHNLRIIMIOYLLSGFELELYSMHEYIYIYWLSEFLYA 532
 QY 540 WLMSTLSRADGSGMAEERIMEEQKGRSSKTKKKKKVRPLSREITMSQAYQNMCAQMPK 599
 Db 533 WLMSTLSRADGSGMAEERIMEEQKGRSSKTKKKKKVRPLSREITMSQAYQNMCAQMPK 592
 QY 600 TWVAEDMDGKVRKPKFELDSQVRYEHRFAPFNSVMTPPVHYLOFKEMSDLNKYSPPQ 659
 Db 593 TWVAEDMDGKVRKPKFELDSQVRYEHRFAPFNSVMTPPVHYLOFKEMSDLNKYSPPQ 652
 QY 660 SPELYVAASKHFQQAAMILENIPNDPHEVNRILKVKAPNFVVMKLLAGHKKESKVPPEF 719
 Db 653 SPELYVAASKHFQQAAMILENIPNDPHEVNRILKVKAPNFVVMKLLAGHKKESKVPPEF 712
 QY 720 DFSAHKYPFVVKLV 733
 Db 713 DFSAHKYPFVVKLV 726

RESULT 4

ABB60388
 ID ABB60388 standard; protein; 784 AA.

AC ABB60388;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 7956.

XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX Drosophila melanogaster.

XX WO2001/71042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL04491.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.

XX Disclosure; SEQ ID NO 7956; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-

CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 784 AA;

QY Query Match 34.5%; Score 1326; DB 4; Length 784;

Best Local Similarity 39.3%; Pred. No. 2.4e-122;

Matches 291; Conservative 131; Mismatches 246; Indels 72; Gaps 14;

QY 29 WVDITQDFEACREIKLGGELHDKLFGLEPMSALEMMDPKMAGMGNQVNRKVLNFRQ 88

Db 80 WVDVTKEFDHCAELQPGELAQDMLEGLFEAMSALEIMDKMDVGVGDFKQDLPPPSFEA 139

QY 89 AIKQGTIKIKDLTLPGLIGIMDTCCCLITWLEHSLAQTVFTCYIINPDIEPAMKA 148

Db 140 AIATGAIKLDLTPSELIGIYDALFSLVSLWEGNSMDQVLTCLYLHAPAIKDKALKRV 199

QY 149 FALGILKICDIAREKVNKAAYFEEDFQSMYTG---FRMANSVTDLRVTGMLKVEDDDMQ 205

Db 200 FCTAVNLIIVIKKIIAIAVAVNEEDFQ--LYGNSALLAAEKAQPATVYSSLKQVDELI 257

QY 206 RRVKSTRSQGEERDPEVELEHQOCLAVFSRVKTRVLLTVLIATFKETKTSABAQA--- 262

Db 258 FKCKKLST-----EDMVAVHRLRFMRHLFQVIYHVEQMASNDTVDDKVDI 304

QY 263 -KLMVQAAALLSAIHNLSHHGIGQANDTTKG-DHPIMMGEFPLVNORLLPPTFPYAKII 320

Db 305 YKILLVASEMLPGIRNTLDRTQPE----KSDAPNPMGFSPRIHDRQPPAFPSIKIR 360

QY 321 KREEMVNYFARLIDRIKTVCEVNLNLHCLIDPFCESEQS-PCVLSRLSLLQTTFLVDN 379

Db 361 DRPSYQFLEBMSIRFKVACKVKYKYVSALNPFIEYSKSGQCILSRSLVQLFSANM 420

QY 380 KKVFTGLMQDMVKDALRSFVDPPVLSPKCYLYNNHQAKDCIDSVTHCV--RPFCSLIQ 437

Db 421 RMAHGKLPKQFLRHSVQVFNPPVLNAKHPVAADPKVQOHLNFRFYCINMNTTQFIR 480

QY 438 IHGHNRRQRDKLGHILEEFATLQDEFMTFFNRAEKVDAALHTMLLKQ-----E 487

Db 481 ICGFNARQRDKLARLIENFTIQ-----VDAARLDSMMQLANERAMEGNE 527

QY 488 PQR---OHLACLGTTWLYHNLRIIMIOYLLSGFELELYSMHEYIYIYWLSEFLYAWLMST 544

Db 528 PMATALKHSHTFSTWVLYNCFRAMLIFLMSGFELELYAVHEFLYIYWPYEFGLFLYSA 587

QY 545 LSRADGSGMAEERIMEEQKGRS-----SKTKKKKKVRPLSREITMSQAYQNM 593

Db 588 LTRTENILLAQEEYAEHQSKTQSGSGGAANKRKAAPKKKKTKQRPYRAEIVFYHALLSL 647

QY 594 CAGMFKTMVAFDMGKVRKPKFELDSQVRYEHRFAPFNSVMTPPVHYLOFKEMSD-LN 652

Db 648 CGMYKAMGALTQDGRVRLPLSKFDNEEIRYNRFLPATLTSPPPVSAYAEFKVREHM 707

QY 653 KSPPPSPPELYVAASKHFQQAAMILENIPNDPHEVNRILKVKAPNFVVMKLLAGHKKKE 712

Db 708 RHS-----VEDLYTYAAKHFDQARNVLESIQNDQEMLDLLQIARTNFVVMVNLARGHOKE 763

QY 713 SKVPPEDFDSAHKYFPVVKLV 732

Db 764 VKRQPEDFDSKHSYFPPIKL 783

RESULT 5

ABG69816

ID ABG69816 standard; protein; 239 AA.

AC ABG69816;

XX 21-OCT-2002 (first entry)

XX Human secretory protein #7.

KW Human; secretory polypeptide; SPTM; actinic keratosis; arteriosclerosis;
KW bursitis; cirrhosis; hepatitis; polycythemia vera; anaemia; psoriasis;
KW primary thrombocytopenia; cancer; adenocarcinoma; leukaemia; myeloma;
KW Parkinson's disease; acquired immunodeficiency syndrome;
KW AIDS; allergy; asthma; Crohn's disease; diabetes mellitus; gout;
KW glomerulonephritis; Goodpasture's syndrome; thyroiditis; pancreatitis;
KW hepatitis; multiple sclerosis; osteoporosis; Reiter's syndrome;
KW rheumatoid arthritis; neurological disorder; epilepsy; stroke; dementia;
KW Alzheimer's disease; Pick's disease; Huntington's disease; mood; anxiety;
KW Parkinson's disease; central nervous system disorder; mental disorder;
KW schizophrenic disorder; amnesia; Tourette's disorder; transgenic animal;
KW gene therapy.
XX Homo sapiens.
XX WO200257304-A2.
XX 25-JUL-2002.
XX 15-JAN-2002; 2002WO-US001340.
XX 16-JAN-2001; 2001US-0261864P.
XX 16-JAN-2001; 2001US-0261865P.
XX 16-JAN-2001; 2001US-0261979P.
XX 16-JAN-2001; 2001US-0261981P.
XX 17-JAN-2001; 2001US-0262164P.
XX 17-JAN-2001; 2001US-0262208P.
XX 17-JAN-2001; 2001US-0263131P.
XX 19-JAN-2001; 2001US-0262599P.
XX 19-JAN-2001; 2001US-0262760P.
XX 19-JAN-2001; 2001US-0263063P.
XX 19-JAN-2001; 2001US-0263066P.
XX 19-JAN-2001; 2001US-0263089P.
XX 19-JAN-2001; 2001US-0263070P.
XX 19-JAN-2001; 2001US-0263074P.
XX 19-JAN-2001; 2001US-0263076P.
XX 19-JAN-2001; 2001US-0263077P.
XX 19-JAN-2001; 2001US-0263329P.
XX (INCY-) INCYTE GENOMICS INC.
XX Panzer SR, Lincoln SE, Altus CM, Dufour GE, Hillman JL, Jones AL;
XX Dam TC, Liu TF, Harris B, Flores V, Daffo A, Marwaha R, Chen AJ;
XX Chang SC, Gerstin EH, Peralta CH, David MH, Lewis SA;
XX WPI; 2002-590716/63.
XX N-PSDB; ABS51309.
XX New purified secretory polypeptides and polynucleotides, useful in the
XX diagnosis, study, prevention or treatment of diseases associated with
XX decreased expression of functional secretory molecules, e.g. AIDS, cancer
XX or allergies.
XX Claim 27; Page 304-305; 340pp; English.
XX PS
XX The invention describes an isolated polynucleotide a naturally occurring
XX polynucleotide sequence at least 90 % identical to it, a polynucleotide
XX complementary to it or an RNA equivalent of it. The purified secretory
XX polypeptides (SPTM) and polynucleotides are useful in the diagnosis,
XX study, prevention or treatment of diseases associated with decreased
XX expression of functional SPTM, e.g. actinic keratosis, arteriosclerosis,
XX bursitis, cirrhosis, hepatitis, polycythemia vera, primary
XX thrombocytopenia, anaemia, psoriasis, cancers including adenocarcinoma,
XX leukaemia, myeloma or sarcoma, immune system disorder such as acquired
XX immunodeficiency syndrome (AIDS), allergies, asthma, Crohn's disease,
XX diabetes mellitus, glomerulonephritis, Goodpasture's syndrome, gout,
XX Hashimoto's thyroiditis, hepatitis, multiple sclerosis, osteoporosis,
XX pancreatitis, Reiter's syndrome, autoimmune thyroiditis or rheumatoid
XX arthritis, neurological disorders such as epilepsy, stroke, Alzheimer's
XX disease, Pick's disease, Huntington's disease, dementia, Parkinson's
XX disease, other developmental disorder of the central nervous system,
XX mental disorder including mood, anxiety or schizophrenic disorder,
XX amnesia or Tourette's disorder. The polynucleotides may be used in

CC hybridisation and amplification technologies, e.g. in assessing gene
CC expression patterns, to develop a transcript image for a particular cell
CC or tissue, or to create transgenic animals to model human disease. This
CC is the amino acid sequence of a human secretory protein isolated in the
CC invention
XX Sequence 239 AA;
SQ
Query Match 28.6%; Score 1100; DB 5; Length 239;
Best Local Similarity 99.5%; Pred. No. 1.2e-100;
Matches 211; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MVMKASVDDDDSGWELSNPKEKSNNTNWVITQDFFEEACRELKLGELHDKLFLGLFEAM 60
Db 28 MVMKASVDDDDSGWELSNPKEKSNNTNWVITQDFFEEACRELKLGELHDKLFLGLFEAM 87
Qy 61 SAITEMDPKMDAGMIGNQVNRKVINFEQAIKDGITIKIKDILPELIGIMDTFCCLITWL 120
Db 88 SAITEMDPKMDAGMIGNQVNRKVINFEQAIKDGITIKIKDILPELIGIMDTFCCLITWL 147
Qy 121 EGHSLAQTVFTCLYIHNPDFTEDPAMKAFALGILKICDIAREKYNKAAVFEEDFQSMY 180
Db 148 EGHSLAQTVFTCLYIHNPDFTEDPAMKAFALGILKICDIAREKYNKAAVFEEDFQSMY 207
Qy 181 GFKMANSVTDLRVTGMLKDVDDMQRRVKSTR 212
Db 208 GFKMANSVTDLRVTGMLKDVDDMQRRVKSTR 239
RESULT 6
ADC32736
ID ADC32736 standard; protein; 160 AA.
XX AC ADC32736;
XX DT 18-DEC-2003 (first entry)
XX DE Human novel contig-encoded polypeptide sequence, SEQ ID NO:2818.
XX KW Human; diagnostic; drug screening; forensics; gene mapping;
KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;
KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
KW ulcers; osteoporosis; autoimmune disease; cancer;
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;
KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnertary;
KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;
KW gene therapy; chromosome 9.
XX OS Homo sapiens.
XX PN WO2003029271-A2.
XX PD 10-APR-2003.
XX PF 24-SEP-2002; 2002WO-US030474.
XX PR 24-SEP-2001; 2001US-0324631P.
XX PA (HYSE-) HYSEQ INC.
XX XX Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
PI Haley-Vicente D, Drmanac RT;
XX WPI; 2003-371981/35.
XX N-PSDB; ADC31969.
XX New polynucleotide and polypeptide useful for diagnosing, preventing or
XX treating conditions such as neurodegenerative diseases, anemias, platelet
XX disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
XX cancer.
XX Example 2; SEQ ID NO 2818; 1185pp; English.
PS

XX The invention relates to 971 novel human cDNA sequences (ADC29919-
 CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The
 CC invention also relates to nucleic acid sequences over 99% identical with
 CC the novel human cDNAs. The invention additionally encompasses expression
 CC vectors and host cells comprising a nucleic acid of the invention; the
 CC recombinant production of a polypeptide of the invention; an antibody
 CC against a polypeptide of the invention; a method of detecting
 CC polynucleotides or polypeptides of the invention; and methods of
 CC identifying a compound which binds to a polypeptide of the invention. The
 CC invention further discloses methods of preventing, treating or
 CC ameliorating a medical condition; kits comprising polynucleotide probes
 CC and/or monoclonal antibodies for carrying out the methods of the
 CC invention; methods for the identification of compounds that modulate the
 CC expression or activity of the polynucleotide and/or polypeptide; and 767
 CC contig sequences corresponding to the cDNA sequences of the invention
 CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
 CC -ADC33394). The nucleic acids and polypeptides of the invention are
 CC useful in diagnostics, drug screening, forensics, gene mapping, in the
 CC identification of mutations responsible for genetic disorders or other
 CC traits, for assessing biodiversity, and in producing many other types of
 CC data and products dependent on DNA and amino acid sequences. They are
 CC also used for treating diseases such as Parkinson's disease, Alzheimer's
 CC disease and other neurodegenerative diseases, anaemia, platelet
 CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
 CC cancer. The nucleic acids may also be used as hybridisation probes or
 CC primers, and in the recombinant production of a protein. The polypeptides
 CC are also useful in generating antibodies, as molecular weight markers,
 CC and as food supplements. The present sequence represents a human contig-
 CC encoded polypeptide sequence used in an example of the invention. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 160 AA;

Query Match 17.8%; Score 685; DB 7; Length 160;
 Best Local Similarity 100.0%; Pred. No. 1.6e-59;
 Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MVMKASVDDDDSGWELSPKWKXSNNTNWVDITQDFEACRELKGLGELLHDKLFLGLPEAM 60
 DB 31 MVMKASVDDDDSGWELSPKWKXSNNTNWVDITQDFEACRELKGLGELLHDKLFLGLPEAM 90
 QY 61 SAIEMDPKMDAGMIGNQVNRKVLNFEQAIKDGTIKIDLPFLPGLIGMDTCFCLITWL 120
 DB 91 SAIEMDPKMDAGMIGNQVNRKVLNFEQAIKDGTIKIDLPFLPGLIGMDTCFCLITWL 150
 QY 121 EGHSLAQTVF 130
 DB 151 EGHSLAQTVF 160

RESULT 7
 AAB07449
 ID AAB07449 standard; protein; 733 AA.

XX AAB07449;

XX 20-OCT-2000 (first entry)

XX Amino acid sequence of a murine tissue repair protein.

XX Tissue repair protein; orofacial clefting; wound healing; tissue repair.

XX Mus sp.

XX WO200040719-A2.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-GB000003.

XX

XX

XX

XX

XX

PR 06-JAN-1999; 99GB-00000167.

XX (UYLE-) UNIV LEEDS.

XX Markham AF, Bonthron D;

XX WFI; 2000-465983/40.

DR N-PSDB; AAA58824.

XX New human and mouse nucleic acids encoding a tissue repair protein,
 PT useful for diagnosing and treating orofacial clefting, and for promoting
 PT wound healing and/or tissue repair.

XX Claim 22; Fig 8; 45pp; English.

XX The present sequence represents a murine tissue repair protein. Tissue
 CC repair gene polynucleotides are useful for determining expression of mRNA
 CC in selected target tissue, e.g. for diagnosing and treating orofacial
 CC clefting. They are also useful for determining the presence of DNA
 CC mutations in patients suffering from, or suspected to be suffering from
 CC orofacial clefting. The antibodies are also useful in the diagnosis of
 CC orofacial clefting. The polynucleotide is also useful for promoting wound
 CC healing and tissue repair

XX Sequence 733 AA;

Query Match 3.3%; Score 127; DB 3; Length 733;

Best Local Similarity 20.9%; Pred. No. 0.0076;

Matches 132; Conservative 91; Mismatches 227; Indels 182; Gaps 31;

QY 217 BERPVEL-----BHQCLAVFSRVKTRVLLTVLIAFTKKTSAVAQAOKLM----- 265

DB 67 EQLDGSLEYDNREHABFVLVRKDLFSQLVETALLALGYSHSSA-AQAQGIILGRWNP 125

QY 266 -----VQADLLSAIHNSLHHGICQANTTKGDHP-----IMWPEPL---- 303

DB 126 LPLSYVDADPATVADMLQDVYHVVTUKIQCSKLEDFABONNHATVNALKELLEKE 185

QY 304 VNQRLPPTFPYAKIKREEMVNYFA-----RLIDRIKTV-CEVVNLNL--H 349

DB 186 MNQSTLAKCEPLSQSMISSIVNVTYANVSATKQCFGRWYKVKIKIVERENLSDY 245

QY 350 CILDFCFEFSQPCVLSR-----SLQTTPLVNKKVFGTHLMQDMVKDALRSFVDPVP 404

DB 246 CVLG-----QRPMLPLPNMNLASLGKTNESPHSQIHSTPIRNV-PALQPTMSPL 297

QY 405 LSPK-----CYLYNN-----HQAKDCID-SFVTH-----CVR-PFCSL 435

DB 298 LSPQLSPQLVRQQTAMAHLINQQTAVSRLLAHQHPQAINQOFLNHPPIPRAVKPEPTNSS 357

QY 436 IQIHGHNARQRDKLGHILBEFATLQDEFMTFFYNRAEKVDAALHTMLLKQEPQOHLAC 495

DB 358 VEVSFDIYQQVRDE-----LKEASVSQAVFARVAFNRQTQ---GLLSEILRKEEDPR----- 405

QY 496 LGTWVLVHLNLRIMQYLLSGFELE---LY-----SMHEYIYIYWLSE----- 535

DB 406 TASQSLLVNLNLRAM-QNFNLNPEVERDRIYQDERERSMNPVSNVSSASSSSPSSRTPOAK 464

QY 536 -----FLYAWLMSTLSRAGDSQMAEERIMEEQQKGRSSKTKKK 574

DB 465 TSPTTDLPIKVDGAVNNTAAIYDEIQCEMKRAKVQALFAKVAANKSQALCWLRLWK 524

QY 575 KKVRLPSREITMSQAYQNCAGMFKTMVAFMDGKVRKPKFELD---SEQVRYEHRFAPF 631

DB 525 ENPSPENRTL-----WENLC-----TIRRF-----LNLQHERDVIYEESRHHSEMRQ 569

QY 632 NSVMTPP-PVHYLQ-----FKMSDLNKYSPPPQSPPELYVAASKHQQAKMILENIP-- 682

DB 570 HVVOLPEPVQVLHROQSQPTKSSPPREAPPPTPTDSCAKKPSRRTKISLEALGIL 629

QY 683 -----NPDHEVNRIL--KVAKPNFVVMK 703

DB 630 QSFTHDVGVPDQEAHHTLSAQDLDPKHTIILK 661

RESULT 8

AAB07448
ID AAB07448 standard; protein; 733 AA.

XX

XX

AC AAB07448;

XX

DT 20-OCT-2000 (first entry)

XX

DE Amino acid sequence of a human tissue repair protein.

XX

XX Tissue repair protein; orofacial clefting; wound healing; tissue repair.

XX

XX Homo sapiens.

XX

XX WO200040719-A2.

FN

XX

PD 13-JUL-2000.

XX

XX 06-JAN-2000; 2000WO-GB000003.

PF

XX

PR 06-JAN-1999; 99GB-00000167.

XX

XX (UYLE-) UNIV LEEDS.

PA

XX Markham AF, Bonthron D;

PI

XX WPI; 2000-465983/40.

DR

XX N-PSDB; AAA58823.

XX

XX New human and mouse nucleic acids encoding a tissue repair protein.

XX

XX useful for diagnosing and treating orofacial clefting, and for promoting

PT

XX wound healing and/or tissue repair.

PT

XX Claim 7; Fig 6; 45pp; English.

XX

XX The present sequence represents a human tissue repair protein. Tissue

XX

XX repair gene polynucleotides are useful for determining expression of mRNA

CC

CC in selected target tissue, e.g. for diagnosing and treating orofacial

CC

CC clefting. They are also useful for determining the presence of DNA

CC

CC mutations in patients suffering from, or suspected to be suffering from

CC

CC orofacial clefting. The antibodies are also useful in the diagnosis of

CC

CC orofacial clefting. The polynucleotide is also useful for promoting wound

CC

XX healing and tissue repair

XX

SQ Sequence 733 AA;

Query Match 3.3%; Score 127; DB 3; Length 733;
Best Local Similarity 20.9%; Pred. No. 0.0076;
Matches 132; Conservative 91; Mismatches 227; Indels 182; Gaps 31;

Qy 217 EERPEVEL-----BHQCLAVFSRVKFRVLLTVLIAFTKXETSAAVAQAQKIM-----265

Dy 67 EQLDGSLEYDNREBAEFVLRKDLVFSQVETALLGALGYSHSA-AQAQGIILKLRWNP 125

Qy 266 -----VQADLLSATNSLHHGIAQNDITKGDHP-----IMWGFEPL----303

Dy 126 LPLSYVTDPADATVADMLQDVVHTLKIQLSCKLEDLFAEQWNHATVRNALKELLKE 185

Qy 304 VNQRLLPPTFRYAKIKREEMVNYFA-----RLIDRIKTV-CEVVNLTLN--H 349

Dy 186 MNQSTLAKCEPLSQSMISVNSTYANVSATKCOEFGWYKYYKIKVERENLSDY 245

Qy 350 CILDFCFEFSQPCVLSR-----SLQTFILVDNKKVFGTHLMQDMVKDALRSFVDPVP 404

Dy 246 CVLG-----QRPWHLPMNQLASLGKTEQSPHSQIHHSPIRNQV-PALQPIMSPOL 297

Qy 405 LSPK-----CYLYNN-----HQAKDCID-SVTH-----CVR--PFCSL 435

Dy 298 LSPQLSPQLVROQIAWAHLNQIAVSRLLAHQHPAINQOFLNHPPIPAVKPEFTNS 357

Qy 436 IQIGHNRRARQDKLGHILIEEPATIQDEPMTTFYFNRAEKVDAAALHTMLLKQEPQRLAC 495

Db 358 VEVSPIYQVRDE---LKRASVSQAVFAFVAFTQ---GLLSEILRKEEDPR-----405
Qy 496 LGTWLVYHNLRIQYLLSGFELE---LY-----SMHEYIYIYWLSE-----535
Dy 406 TASQSLVNLRLAM-QNFINLPFEVERDRIYQDERSRMNPVMSVSSASSPSSRTPOAK 464
Qy 536 -----FLYAWLMSTLSRADGSGMAEERIMBERQKGRSSKTKKK 574
Dy 465 TSTPTDLPKVDGANINITAAIYDEIQEEMKRAKVSQALFAKVAANKSQWLCELLRWK 524
Qy 575 KKVRLPSREITMSQAYQNMCMGMEKFTWAFDMGKVRKPKFELD---SEQVRYHRFAPF 631
Dy 525 ENPSPENRTL-----WENLC-----TLRRP-----LNLQHERDVIYEESRHHSEMRM 569
Qy 632 NSVMTTPP-FVHYLO-----FKEMSDLNKYSPPSQSPELYVAASKHFQQAQKMLENP--682
Dy 570 HVQLPPEPVQVLHROQSQFAKSSPPREEAPPPPTEDSCAKKPRSRTKISLEALGIL 629
Qy 683 -----NPDHEVNRIL--KVAKPNEFVVMK 703
Dy 630 QSFHVDGLYDPQEAHTLSAQDLDPKHTIIRK 661

RESULT 9

AAE36464

ID AAE36464 standard; protein; 761 AA.

XX

AC AAE36464;

XX

DT 07-AUG-2003 (first entry)

DE Human colon tumour protein from clone, C1562P KIAA1034.

XX

XX Human; colon tumour protein; cancer; gene therapy; vaccine.

XX

XX Homo sapiens.

XX

XX WO2003022126-A2.

XX

XX 20-MAR-2003.

XX

XX 20-AUG-2002; 2002WO-US026601.

XX

XX 21-AUG-2001; 2001US-0314221P.

XX

XX 21-DEC-2001; 2001US-0343517P.

XX

XX (CORI-) CORIXA CORP.

XX

XX Meagher MJ, King GE, Secrist H, Jiang Y, Stolk JA;

XX

XX WPI; 2003-300990/29.

XX

XX N-PSDB; AAD55553.

XX

XX New isolated polynucleotide useful for stimulating and/or expanding T-

XX

XX cells specific for a tumor protein, and preventing, diagnosing and/or

XX

XX treating cancer, particularly colon cancer, in patients.

XX

XX Claim 2; Page 146-148; 158pp; English.

XX

XX The invention relates to compositions and methods for the diagnosis and

XX

XX therapy of cancer, particularly colon cancer. The invention also provides

XX

XX colon tumour proteins and to polynucleotides encoding such proteins.

XX

XX Polynucleotides of the invention are useful for preventing, diagnosing

XX

XX and/or treating cancer, particularly colon cancer in a patient. They are

XX

XX used in gene therapy and also to prepare vaccines. The present sequence

XX

XX is human colon tumour protein

XX

SQ Sequence 761 AA;

Query Match 3.3%; Score 127; DB 6; Length 761;
Best Local Similarity 20.9%; Pred. No. 0.008;
Matches 132; Conservative 91; Mismatches 227; Indels 182; Gaps 31;

Db 8 RELBECKSLR-----SRRNQLEERSVAME--DEMNMKREGKPREKRIKNEQSLQEI 59
Qy 94 TIKIKDITLPELIGIMDFCFCLITWLBGHSIAQVFTCLYIHNPDFIEDPAMKAFALGI 153
Db 60 WDVYKRNPL-RLIGVPES-----DAENGTKLENTL-----QDIIQEDFPNLRQAN 104
Qy 154 LKICDIAR-----EKVNAKAAVEEDFQS----- 177
Db 105 VQIQEIQTPOKYSRRATPRHIIVRFTKVMKQKMLRAA--REKDFKPTKIKRDKEGHY 162
Qy 178 -MTVGFKNANSVTDLRV---TG---MLKDYEDDMQRRVKSTRSQGERDPVELEHQ 229
Db 163 IMVKGSIQOEELTILNIYAPNTGAPRFIKQVLSLDLQDLSHTLIMGDFNTP----- 214
Qy 230 CLAVFSRVKFTRVLLTVLIAFTKETSVAEAAQKLMVQAADLLSAIHNSLHHG-----I 283
Db 215 -----LSTLDRSTRQKVNKDTQELNSALHQAALID-IYRTLPKSTYEYTF 259
Qy 284 QAQNDT-TKGDPHIMMGPEPLVNORLLPPTFPFYAKIIKREEMVNYFARLIDRINKTUCEV 342
Db 260 SAPHTYSKIDH--IVGSKALLSK-----CKRTQIITNYSLSHSAIKLELRI 304
Qy 343 VNLNHLHCILDFCFEFSQSVLSRLQTTFLVDNKKVGTGTHLMQDMVKDALRSFVDP 402
Db 305 KNLQ-----SR-----TTWKLNNMLNDYVWVHNKMAEIKMFFET 342
Qy 403 PVLSPKCYLYNHOAKDCIDSVTHCVRPFCSLIQIHGHNARQDKLGH----- 453
Db 343 KENKDTTY-----QNLWDAFAKVRCKF---IALNAHKRQERSKIDTILTSQKLEK 392
Qy 454 -----LEFATIQ--DEFMTFVFNRAEKVDAALHTMLLKOE 487
Db 393 QEOTNHSKASRROBITKIRABKEIETQTLQKINESRWFPERINKIDRPLARLIKRR 452
Qy 488 POROHLACLTGWLYHNLRIMIQVLSG-----FELEYSMEHYIYIYWLSEFLYA 539
Db 453 EKNQ-----IDTIKNDKGDITDPTDTEIQT-TIREYY-----KHLYA 487
Qy 540 WLMSTLRADGSGMAEERIMEEQKGRSSKTKKKKVRPLSRITMSQAYQNMCAGMFK 599
Db 488 NKLENLEMD-----KFLDTYTLPLRN-----QEEVESLNRPTDSEI--EALINSLP 533
Qy 600 TWAFDMGDKVRKPKFELDSQVRYEHRFAPFNSVMTPPVHYLQFKEMSLINKYPPPP 659
Db 534 TKKSPGLD-----FTAELYQ-RYKEELVFP-----LLKLFQSIKIERILPN 574
Qy 660 SPELYVAASKHFQAKMILENIPNPDHEVNRILKVAKENFVVMKLL 705
Db 575 S-----FHEASIIIL--IPKGRDTAK-----KENFRPISLM 603

RESULT 15
ABG02338
ID ABG02338 standard; protein; 1184 AA.
XX
XX
AC ABG02338;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #2329.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
ED
XX 30-MAR-2001; 2001WO-US008631.
PF
XX
XX 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS66525.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 32697; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1184 AA;

Query Match 3.2%; Score 121.5; DB 4; Length 1184;
Best Local Similarity 20.6%; Pred. No. 0.057;
Matches 123; Conservative 87; Mismatches 201; Indels 187; Gaps 31;

Qy 171 BEEDFQSMYTGFKMANSVTDLRV---TG---MLKDYEDDMQRRVKSTRSQGERDDEV 223
Db 82 DKEGHYLMVKGSIQOEBELTILNIYAPNTGAPRFIKQVLTDLQDLSHTLIMGDNTP-- 139
Qy 224 ELEHQOCLAVFSRVKFTRVLLTVLIAFTKETSVAEAAQKLMVQAADLLSAIHNSLHHG- 282
Db 140 -----LSTLDRSTRQKVNKDTQELNSALHQAALID-IYRTLPK 178
Qy 283 -----IQAQNDT-TKGDPHIMMGPEPLVNORLLPPTFPFYAKIIKREEMVNYFARLIDRI 336
Db 179 TEYTFPSAPHTYSKIDH--ILGSKALLSK-----CORTEIITNYSLSHSAI 223
Qy 337 KTVCEVNLTLNHLHCILDFCFEFSQSVLSRLQTTFLVDN-----KKVFGTHLMQD 390
Db 224 KLELRINKLITQNH-----STTWKLNNLLNDYVWVHNEMKAIKMFPEINENKD 271
Qy 391 M-----VKDALRSFVDPVLSPKCYLYNNH-----QAKDCIDSVTHCVRPFCSLIQIHG-N 442
Db 272 TTYQNLMDAFKA-----VCRGKFVALNAHKRQERSKIDT-LTSQKLEKEQEQTSHKAS 325
Qy 443 RARQDKLGHILEEF---ATLQ--DEFMTFVFNRAEKVDAALHTMLLKOEPRQHLACLG 497
Db 326 RROQITKIGAELEIETQTLQKINESRWFPERINKIDRPLARLIKRRKNO----- 379
Qy 498 TWLYHNLRIMIQVLSG-----FELEYSMEHYIYIYWLSEFLYANLMTLSRAD 549
Db 380 -----IDAINKDKGDIITDPTDTEIQT-TIREYY-----KHLYANKLENL---- 416
Qy 550 GSQMAEERIMEEQKGRSSKTKK--KKVFPLESEITMSQAYQNMCAGMFKTWAFDMD 607


```
Db 417 -----REMDFLNTYTLPRLNQEEVESLNRPTGSEIV--AIINSLRTKSPGPD 464
Qy 608 GKVRKPKFELDSEQVRVEHREAPFNSVMTPPVHYLQFKEMSDLNKYSPPPQSPELYVAA 667
Db 465 G-----FTAEFYQ-RYKEELVPF-----LLKQFQSTIEKEGILFNS----- 498
Qy 668 SKHFQQAAMILENIPNDHEVNRILKVAKPNF-----VVMKLLAGGHKKESK 714
Db 499 ---FYEASILL--IPKGRDTTK-----KENFRPMSLMNIDAKILNKILANGIOOHIX 546
```

Search completed: August 10, 2004, 21:50:11
Job time : 63 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 21:52:08 ; Search time 50 Seconds
(without alignments)
4598.592 Million cell updates/sec

Title: US-10-001-857-145

Perfect score: 3849

Sequence: 1 MVMKASVDDDDSGWELSMPE.....KVPPEFDSAHKYPVVVKLV 733

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3849	100.0	733	13	US-10-001-857-145
2	3783	98.3	725	16	US-10-467-433-3
3	595.5	15.5	706	16	US-10-437-963-146818
4	322.5	8.4	256	12	US-10-424-599-190465
5	206.5	5.4	355	12	US-10-424-599-220965
6	139.5	3.6	965	12	US-10-403-571-40
7	136	3.5	2419	16	US-10-408-765A-2722
8	127	3.3	761	14	US-10-225-486-55
9	127	3.3	761	16	US-10-408-765A-696
10	127	3.3	969	12	US-10-335-977-7409
11	127	3.3	972	12	US-10-335-977-7410
12	121	3.1	745	12	US-10-425-114-65126
13	119	3.1	3353	9	US-09-888-615-64
14	116.5	3.0	1441	15	US-10-412-897-3
15	116.5	3.0	1641	12	US-09-964-956-40
					Sequence 145, App
					Sequence 3, Appli
					Sequence 146818,
					Sequence 190465,
					Sequence 220965,
					Sequence 40, Appl
					Sequence 2722, Ap
					Sequence 55, Appl
					Sequence 696, App
					Sequence 7409, App
					Sequence 7410, App
					Sequence 65126, A
					Sequence 64, Appl
					Sequence 3, Appli
					Sequence 40, Appl

```

16 116.5 3.0 1641 13 US-10-017-216-5
17 116 3.0 736 12 US-10-170-385-285
18 115 3.0 633 14 US-10-097-340-23
19 115 3.0 1187 13 US-10-002-769-14
20 115 3.0 1187 13 US-10-024-623-34
21 115 3.0 1187 14 US-10-154-419-46
22 115 3.0 1187 14 US-10-154-419-84
23 115 3.0 1187 14 US-10-146-733-79
24 114.5 3.0 892 16 US-10-437-963-200109
25 114.5 3.0 1191 15 US-10-161-493-52
26 114.5 3.0 1603 15 US-10-369-493-6744
27 114 3.0 1679 15 US-10-369-493-22080
28 113.5 2.9 917 16 US-10-437-963-200112
29 113.5 2.9 1192 12 US-10-332-447-17
30 113.5 2.9 1957 15 US-10-369-493-2070
31 112 2.9 679 16 US-10-437-963-175526
32 112 2.9 1012 16 US-10-437-963-110215
33 112 2.9 1786 9 US-09-742-096-3
34 112 2.9 1787 12 US-10-415-253-2
35 111.5 2.9 1332 13 US-10-041-856-3
36 111.5 2.9 1332 13 US-10-041-856-5
37 111 2.9 662 12 US-10-424-599-164683
38 110 2.9 909 13 US-10-078-929-168
39 110 2.9 909 14 US-10-078-770-190
40 109.5 2.8 533 12 US-10-282-122A-61253
41 109.5 2.8 1321 16 US-10-437-963-191477
42 109.5 2.8 1401 16 US-10-287-226-142
43 109.5 2.8 2427 15 US-10-369-493-6734
44 109 2.8 794 15 US-10-320-797-3294
45 109 2.8 981 9 US-09-815-242-12211

```

ALIGNMENTS

RESULT 1

```

US-10-001-857-145
; Sequence 145, Application US/10001857
; Publication No US20020183500A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prote
; FILE REFERENCE: DEX-0273
; CURRENT APPLICATION NUMBER: US/10/001.857
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,054
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 145
; LENGTH: 733
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-857-145

```

```

Query Match 100.0%; Score 3849; DB 13; Length 733;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 733; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVMKASVDDDDSGWELSMPEKSNNTNWDTQDFEACRELKGLHLDKFLGLFEAM 60
Db 1 MVMKASVDDDDSGWELSMPEKSNNTNWDTQDFEACRELKGLHLDKFLGLFEAM 60
Qy 61 SAIEIMPDKMDAGMIGNQVNRKVLNFQAIKDGTIKIKDLTLPGLIGIMDTFCCLITWL 120
Db 61 SAIEIMPDKMDAGMIGNQVNRKVLNFQAIKDGTIKIKDLTLPGLIGIMDTFCCLITWL 120
Qy 121 EGHSLAQVTFTCLYIHNPDFIEDPAMKAFALGILKICDIAREKVNKAATVEEDFQSMY 180

```

Db 121 EGHSLAQVTFCLYIHNPDIEDPAMKAFALGILKICDIAREKVNKAAREEEDFQSMY 180
QY 181 GFKWANSVTDLRVTGMLKVEDDMQREVKSTRSQGEERPEVELEHQOCLAVFSRVKFT 240
Db 181 GFKWANSVTDLRVTGMLKVEDDMQREVKSTRSQGEERPEVELEHQOCLAVFSRVKFT 240
QY 241 RVLLTVLIAFTKKTSAVAEAKLMVQAADLLSAIHNSLHHGIGQAQNDTTKGDHPIMMGF 300
Db 241 RVLLTVLIAFTKKTSAVAEAKLMVQAADLLSAIHNSLHHGIGQAQNDTTKGDHPIMMGF 300
QY 301 EPLVNQRLPPTPRYAKIKREEMVNYFARLIDRIKTVCVVNLTLNHLICLDPFCFSE 360
Db 301 EPLVNQRLPPTPRYAKIKREEMVNYFARLIDRIKTVCVVNLTLNHLICLDPFCFSE 360
QY 361 QSPCVLSRSLQTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVPLSPKCYLYNNHQAKDC 420
Db 361 QSPCVLSRSLQTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVPLSPKCYLYNNHQAKDC 420
QY 421 IDSFVTHCVRPFCSLIQIHGHNARQDKLGHILEEPATLQDEPMTFFNRAEKVDAALH 480
Db 421 IDSFVTHCVRPFCSLIQIHGHNARQDKLGHILEEPATLQDEPMTFFNRAEKVDAALH 480
QY 481 TMLLKQBPQOHLACLGTWVLYHNLRIMIQYLLSGFELELYSMHEYIYIYWLSEFLYAW 540
Db 481 TMLLKQBPQOHLACLGTWVLYHNLRIMIQYLLSGFELELYSMHEYIYIYWLSEFLYAW 540
QY 541 LMSTLSRADGSQMAEERIMEEQKGRSSKTKKKKKVRPLSREITMSQAYQNMCMGFKT 600
Db 541 LMSTLSRADGSQMAEERIMEEQKGRSSKTKKKKKVRPLSREITMSQAYQNMCMGFKT 600
QY 601 MVAFDMGDKVRKPKFELDSQVRYEHRFAPNSVMTPPVHYLQFKEMSLNKYSPPPQS 660
Db 601 MVAFDMGDKVRKPKFELDSQVRYEHRFAPNSVMTPPVHYLQFKEMSLNKYSPPPQS 660
QY 661 PELVVAASKHFQOAKMILENIPNPDHEVNRILKVAKNFVVMKLLAGGHKESKVPPEFD 720
Db 661 PELVVAASKHFQOAKMILENIPNPDHEVNRILKVAKNFVVMKLLAGGHKESKVPPEFD 720
QY 721 FSAHKYFPVVKLV 733
Db 721 FSAHKYFPVVKLV 733

RESULT 2

US-10-467-433-3
; Sequence 3, Application US/10467433
; Publication No. US2004008773A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION;
; APPLICANT: LAL, Preeti G.; BAUGHN, Mariah R.;
; APPLICANT: YAO, Monique G.; CHAWLA, Narinder K.;
; APPLICANT: ELLIOTT, Vicki S.; XU, Yuming;
; APPLICANT: HONCHELL, Cynthia D.; YUE, Henry;
; APPLICANT: DING, Li; GIETZEN, Kimberly J.;
; APPLICANT: ISON, Craig H.; LU, Dyung Aina M.;
; APPLICANT: HAFALIA, April J.A.; GANDHI, Ameena R.;
; APPLICANT: THANGAVELU, Kavitha; SANJANWALA, Madhusudan M.;
; APPLICANT: TANG, Y. Tom; RAMKUMAR, Javalaxmi;
; APPLICANT: GRIFFIN, Jennifer A.; SWARNAKAR, Anita;
; APPLICANT: AZIMZAI, Yajda; SAPPERSTEIN, Stephanie K.;
; APPLICANT: BURFOZ, Neil; LEE, Ernestine A.;
; APPLICANT: LU, Yan; TRAN, Uyen K.;
; APPLICANT: MARQUIS, Joseph P.
; TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT
; FILE REFERENCE: PP-0899 USN
; CURRENT APPLICATION NUMBER: US/10/467,433
; CURRENT FILING DATE: 2003-08-06
; PRIOR APPLICATION NUMBER: PCT/US02/03709
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/268,117
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: US 60/269,618
; PRIOR FILING DATE: 2001-02-15

; PRIOR APPLICATION NUMBER: US 60/271,118
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: US 60/274,486
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 60/274,436
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/334,229
; PRIOR FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: US 60/353,284
; PRIOR FILING DATE: 2002-02-01
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PERL Program
; SEQ ID NO 3
; LENGTH: 725
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 7111920CD1
US-10-467-433-3

Query Match 98.3%; Score 3783; DB 16; Length 725;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 724; Conservative 0; Mismatches 1; Indels 8; Gaps 1;
QY 1 MVKASVDDDDSGWELSMPEKSNNTNWVDITQDPFEACRELKLGELLHDKLGLFEAM 60
Db 1 MVKASVDDDDSGWELSMPEKSNNTNWVDITQDPFEACRELKLGELLHDKLGLFEAM 60
QY 61 SAETMDPKMDAGMIGNQVNRKVLNFEQAIKDGTIKIKDITLPELIGIMDTFCCLITWL 120
Db 61 SAETMDPKMDAGMIGNQVNRKVLNFEQAIKDGTIKIKDITLPELIGIMDTFCCLITWL 120
QY 121 EGHSLAQVTFCLYIHNPDIEDPAMKAFALGILKICDIAREKVNKAAREEEDFQSMY 180
Db 121 EGHSLAQVTFCLYIHNPDIEDPAMKAFALGILKICDIAREKVNKAAREEEDFQSMY 180
QY 181 GFKWANSVTDLRVTGMLKVEDDMQREVKSTRSQGEERPEVELEHQOCLAVFSRVKFT 240
Db 181 GFKWANSVTDLRVTGMLKVEDDMQREVKSTRSQGEERPEVELEHQOCLAVFSRVKFT 240
QY 241 RVLLTVLIAFTKKTSAVAEAKLMVQAADLLSAIHNSLHHGIGQAQNDTTKGDHPIMMGF 300
Db 241 RVLLTVLIAFTKKTSAVAEAKLMVQAADLLSAIHNSLHHGIGQAQNDTTKGDHPIMMGF 300
QY 301 EPLVNQRLPPTPRYAKIKREEMVNYFARLIDRIKTVCVVNLTLNHLICLDPFCFSE 360
Db 301 EPLVNQRLPPTPRYAKIKREEMVNYFARLIDRIKTVCVVNLTLNHLICLDPFCFSE 360
QY 361 QSPCVLSRSLQTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVPLSPKCYLYNNHQAKDC 420
Db 361 QSPCVLSRSLQTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVPLSPKCYLYNNHQAKDC 420
QY 421 IDSFVTHCVRPFCSLIQIHGHNARQDKLGHILEEPATLQDEPMTFFNRAEKVDAALH 480
Db 421 IDSFVTHCVRPFCSLIQIHGHNARQDKLGHILEEPATLQDEPMTFFNRAEKVDAALH 480
QY 481 TMLLKQBPQOHLACLGTWVLYHNLRIMIQYLLSGFELELYSMHEYIYIYWLSEFLYAW 540
Db 473 TMLLKQBPQOHLACLGTWVLYHNLRIMIQYLLSGFELELYSMHEYIYIYWLSEFLYAW 532
QY 541 LMSTLSRADGSQMAEERIMEEQKGRSSKTKKKKKVRPLSREITMSQAYQNMCMGFKT 600
Db 533 LMSTLSRADGSQMAEERIMEEQKGRSSKTKKKKKVRPLSREITMSQAYQNMCMGFKT 592
QY 601 MVAFDMGDKVRKPKFELDSQVRYEHRFAPNSVMTPPVHYLQFKEMSLNKYSPPPQS 660
Db 593 MVAFDMGDKVRKPKFELDSQVRYEHRFAPNSVMTPPVHYLQFKEMSLNKYSPPPQS 652
QY 661 PELVVAASKHFQOAKMILENIPNPDHEVNRILKVAKNFVVMKLLAGGHKESKVPPEFD 720
Db 653 PELVVAASKHFQOAKMILENIPNPDHEVNRILKVAKNFVVMKLLAGGHKESKVPPEFD 712

QY	721	FSAHKYPVPVKLV	733
Db	713	FSAHKYPVPVKLV	725

RESULT 3

US-10-437-963-146818

Sequence 146818, Application US/10437963

Publication No. US2004012343A1

GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J.

APPLICANT: Kovalic, David K.

APPLICANT: Zhou, Yihua

APPLICANT: Cao, Yongwei

APPLICANT: Wu, Wei

APPLICANT: Boukharov, Andrey A.

APPLICANT: Barbazuk, Brad

APPLICANT: Li, Ping

TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

FILE REFERENCE: 38-21(53221)B

CURRENT APPLICATION NUMBER: US/10/437,963

CURRENT FILING DATE: 2003-05-14

NUMBER OF SEQ ID NOS: 204966

SEQ ID NO 146818

LENGTH: 706

TYPE: PRT

ORGANISM: Oryza sativa

FEATURE:

NAME/KEY: unsure

LOCATION: (1)..(706)

OTHER INFORMATION: unsure at all Xaa locations

OTHER INFORMATION: Clone ID: PAT_MRT4530_47405C.1.pep

US-10-437-963-146818

Query Match 15.5%; Score 595.5; DB 16; Length 706;

Best Local Similarity 25.4%; Pred. No. 6.4e-48;

Matches 191; Conservative 138; Mismatches 296; Indels 135; Gaps 25;

QY	25	SNTNWVDITQDFEACRELKGLGELHDKLGLFEAMSAIEMMDPKMDAGMIGNQVNRKVL	84
Db	21	ASSAWADAPLLDAACRDLQDGLIHGENSLFGLFAMSALEIMDPKMDCGI----	EKSYY 76
QY	85	NFEQAIKDGTIKI-----KDLTLPGLIGIMDTFCCLITWLEGHSLAQTVFTCLYIHNPD	139
Db	77	SIDEAIEDGIAPVLSLDRLTDIQLDVMDFHLSCEATWHKGTLLAQTVFTCLYLMRME	136
QY	140	FIEDPA-WKAFALGILKCDIAREKNKAAVFEEDFQSMYTFKMANSVTDLRVTGMLK	198
Db	137	RTSSHALLNSFCRILRATCAVAVSWVSTARHEDLFTMSFGLPLRDE-GDEKCLSLN	195
QY	199	DVEDDMORRVKSTSR-----QG-EERDPVELEHQCLAVFSRVKTRVLLITVLLIA	249
Db	196	SVEETIARQURACAQALSRKKTLEGLSLQDNPDLEEDYCRALLCLRFRKHFHYVTC	255
QY	250	FTKETSAAVEAQAQ-----LMVQAADLLSAIHNLSLHGGIAQND-----TTKGDHP	295
Db	256	LRPHGRLGLAQHVASCUTLSLMSLRDFLSQSN-----TQOGDEICTTASGRP	311
QY	296	IMMGFEELVNQRLLPPTFRYAKIIKREWMYFARLIDRIKTVCEVNVNTNLHCILDF	355
Db	312	V--GFDASLNSRLSPAPRAVKLSWSDAIRYPEKLLRDLIDICSSPLDPLVLENLHFV	369
QY	356	CFSEQSPCVLSRLQTTFLVDNKKVFGTHLMDQMKDALRSFDPVPLSPKCYLNNH	415
Db	370	VQFQKSPDLVPRAFLO-TLLVQDGKLYGRDLSCDVSRALS-----PDIIGDEQFQMN--	423
QY	416	QAKCIDSFVTHCVRPFCCLIQIHGHNARORDKLGHLLEEFATLQDEMTFYENR--AE	473
Db	424	-----EFVVQLGQLVNLKULICTNFAWRRKLGKSLQDWSTISIQ-LEFALKREFGE	475
QY	474	KVDAALH-TMLLKQEPQRQHLACLGTVLVLNLRIMTIQYLLSGFELELYSNHYYITYWY	532

Db	476	TRNVLPHENNCMRVSKQ-----LIVWTOEHYVWAYRELILGFLDLYSPSEYCMVY--	527
QY	533	LSEFLYAWLWMLSTLSRADGSGMAEERIMEEQOQGRSSKTKKKKKVPLSREIT-----M	586
Db	528	-----XRRKGK-----KKDHSKDSRDTAFPSCLL	554
QY	587	SQAYQNMCAGM-FKTMVAFDMGKVR---KPKFELDSEQVRYEHRFAPFNSVMTPPPVHY	642
Db	555	LQCYVLLSEGLSMDFOAFAFSQXIHINSHWNMKLEFTALRGRTQHFLLQKARLPEHITY	614
QY	643	LQFKE-----MSDLNKYSPFPQSPELYVAASKHFQOAKMILENI-----PNPDHVN	689
Db	615	YSFRESASHASIAADLTKN-----FFKEIHKITPSLRGSPASEPEKLAIEIR	660
QY	690	RILKVAKPFPVVMKLLAGGHKXESKVPPEDFDSAHKYFPV	729
Db	661	QIEQVAEHRNRIALNIISQVAGDPSLRVSFEFTHHPHFAV	700

RESULT 4

US-10-424-599-190465

Sequence 190465, Application US/10424599

Publication No. US20040031072A1

GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J

APPLICANT: Kovalic, David K

APPLICANT: Zhou, Yihua

APPLICANT: Cao, Yongwei

TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

FILE REFERENCE: 38-21(53223)B

CURRENT APPLICATION NUMBER: US/10/424,599

CURRENT FILING DATE: 2003-04-28

NUMBER OF SEQ ID NOS: 285684

SEQ ID NO 190465

LENGTH: 256

TYPE: PRT

ORGANISM: Glycine max

FEATURE:

NAME/KEY: unsure

LOCATION: (1)..(256)

OTHER INFORMATION: unsure at all Xaa locations

OTHER INFORMATION: Clone ID: PAT_MRT3847_143009C.1.pep

US-10-424-599-190465

Query Match 8.4%; Score 322.5; DB 12; Length 256;

Best Local Similarity 31.8%; Pred. No. 3.2e-22;

Matches 77; Conservative 49; Mismatches 93; Indels 23; Gaps 6;

QY	17	SMPEKM--EKSNTNWVDITQDFEACRELKGLGELHDKLGLFEAMSAIEMMDPKMDAG	73
Db	6	SLPRAISFADNSWADVSPDLLQAACQDQGLIHGDNFNLFAMSALEIMDPKMDSG	65
QY	74	MIGNQVNRKVLNFEQAIKDGTIKI-----KDLTLPGLIGIMDTFCCLITWLEGHSLAQ	128
Db	66	-----IACYTSLDEALENGVAPVPSADKTTDVRCMIDIMDHLACEATWHKGHSLAQ	120
QY	129	VFTCLYIHNPDFTIEDPA-WKAFALGILKCDIAREKNKAAVFEEDFQSMYTFKMAN	187
Db	121	VYSCLYLRPERTSHALLSHYKVRATCAKLSVVSQARTHEEDLFTMAYGLPLSGD	180
QY	188	VTDLRVTGMLKQVEDDMQRRVKSTRSQGEERDPE-----VELEHQCLAVFSRVK	239
Db	181	-GDEKCLSLNLAIVEETISQLRACKVSSSKRRVPDIEPLQNNPDLEEGYCKALXCLRF	239
QY	240	TR 241	
Db	240	RK 241	

RESULT 5


```
Qy 645 -----FKMSDLN-----KYSPPQSPPELVAAASKHQQAAMILENI--- 681
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 779 DTVIAIILKIPLSFORFYFQKLQSTSIKIALSPSPRTNEPIPVQNN-QQLTLKVEGVIQH 837
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 682 ---PNPDHEVNR1-LKVA-----KP--NFVVMKLLAGSHKESKVPPEFDSAHY 726
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 838 GSSPGLPRKIQAVCLKVSSTLTQPGSDFKI-PLESKTNEIEQKRVPHNDYFSTQF 892
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
US-10-408-765A-2722
; Sequence 2722, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fany, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Marnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2722
; LENGTH: 2419
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2722

Query Match 3.5%; Score 136; DB 16; Length 2419;
Best Local Similarity 19.4%; Pred. No. 0.012;
Matches 176; Conservative 141; Mismatches 303; Indels 288; Gaps 49;

Qy 13 GWELSMPEKMEKSNW---VDITQD-----PEEACRELKGLGELHDKLFGIFEAMSALEM 65
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1573 GNEEAMKEQLEQLKEHWDHLLERTNDKGLNEASRQORFNTSIRDFFELWSEAFETLLAM 1632
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 66 MDPKMDAGMIGN-----QVNRKVLNPEQATKDGITIKILTLPGLIGIMDTFCCLITW 119
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1633 KQQRDLASAGNLKXQLLERELAREDAKDINTLAEDL----- 1673
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 120 LEHSLAQTVFTCLYIHNPDFI---EDPAMKAFALGILKICDIAREKVNKA-AVF----- 170
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1674 -----LSSGTF-----NVDQIVKKKDNVNRKF-LNVQELAAAHHEKLKEAYALFQFPQD 1721
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 171 --BEEDF-----QSMYGFKNVSVTDL-----RVTGMLKQVED-----DMQRVK 209
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1722 LDDEESTIEKLIIVSQDYG-RDLQGVQNLKXKHLRLEGELVAHEPAIQNVLDMAEKLK 1780
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 210 STRQRQERDPEVEL-----EHQOCLAVFSRVKFTVLLTV-----LIAFTKETSAVA 259
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1781 D-KAAGVQE---ETQLRLAQFVEHWEKLEKARGLKLBSLEVLQPMQNAEBEAWIN 1836
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 260 EAQKLMVQ--AADLLISAI-----HNSLH-----HGIQAQNDTTKGHPIMMGFEPLVQ 306
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1837 EKNALAVRGDCGTAAATQSLMLKHEALENDFAVHETRVQNVCAQG-----EDILN- 1887
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 307 RLLPPTFPFYAKIIKRE-----EMVNYFARLIDRIKTVCVVNLTNLHCILDF-FCERSEQ 361
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1888 -----KVIQESQNKKEISKIEALNEKFTSLAKAAMKQLQLEDDYAFQEFNWK 1936
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 362 SPCVLSRLQTTFLVADNKKVFGTHL-----MQDMVKDALRSFVD---PPVLSPKCY 410
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1937 ADVVEAWIADKETSILKTGN--GADLGDFTLLAKQDTILDASLQSFQOERLPEITDLKDK 1994
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 411 LYN-NHQAKDCIDSPVTHCVHCPFSLSLIQIHGHNRRARQDKLGHILEEPATIQ---DEFMT 466
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1995 LISAGHNQSKAIEERYAALLKRWELLESASAVHRQK-----LLEKQLPLQKAEIDLFE 2047
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Qy 467 FYF-----NRAEKVDAAL-----HTMLLKQ--EPORQH----- 492
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2048 FAHKASALNNWCXEENLSEPVHCVSLNEIRLOKDHEDFLASLARAQADFCKLLELDQ 2107
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 493 -LACLG-----TWVLYHNLRIMIQYL-----LSGFEL-ELYSMHEY 526
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2108 QIKALGVSPSPYTWLTVVEVLERTWKHLSDIIEBEQELQKEARQVKNFECQFEQNAS 2167
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 527 YIYVYLSSEFLYAWLMSTLSRADGQSWABERIMBEQOGRSKSKTKKKKKVRLPSREIT- 585
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2168 TFLQWILETRAYF-----LDGSLKKTGTLSQLEANK-----RKQKETQAMKRQLTK 2215
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 586 -----MSQAYQ-----NWCAGMFKTMVAFDMGKVRKPKFE 616
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2216 IVDLGDNLDELALIDIKYSTIGLAQQWDQYQLGRMOHLEQOIQAKDINGKVSEETLKE 2275
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 617 LDSQVRYEHRFAFPNSVMTPTPPVHYLQFKE-MSDLNKKYSPPPQSPEL----- 663
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2276 FST---LYKH-----FDENLTGRLTH-KEFRSCLRGLNVLPMVEEDEHEPKFEKFLDAVD 2327
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 664 -----YVAASKHFQQAAMI---LENIPNPDHEVNRILKVAKPVPVVKLLAGGHKESKV 715
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2328 PGRKGVVSLDEY--TAFLLDKESENIKSSDEIENAFQALAEKSYITK-----EDMKQALT 2381
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 716 PPEFDFA 723
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2382 PEQVSFCA 2389
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
US-10-225-486-55
; Sequence 55, Application US/10225486
; Publication No. US20030129207A1
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: King, Gordon E.
; APPLICANT: Secrist, Heather
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.591
; CURRENT APPLICATION NUMBER: US/10/225,486
; CURRENT FILING DATE: 2002-10-29
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 55
; LENGTH: 761
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-225-486-55

Query Match 3.3%; Score 127; DB 14; Length 761;
Best Local Similarity 20.9%; Pred. No. 0.014;
Matches 132; Conservative 91; Mismatches 227; Indels 182; Gaps 31;

Qy 217 BEEDPEVEL-----BHQOCLAVFSRVKFTVLLTVLIAFTKETSASVAEAQKLM----- 265
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 95 EQLDGSLEYDNREBHAEBFLVRKDLVLSQLVETALLALGYSHSSA-AQAQGIILKGRWP 153
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 266 -----VQAADLLSAIHNSLHHGICQANDTTKGDHP-----IMMGFEPL- 303
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 154 LPLSVYTDADPATVADMQLQDVYHVVTTLKIQLOCSKLEDLPAEQWNHATVNAKLELKE 213
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 304 VNQRLLPPTFPFYAKIIKREEMVNYFA-----RLIDRIKTV-CEVYNLTNL--H 349
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 214 MNQSTLAKCEPLSQSMISSIVNSYIYANVSATKQCFGRWYKKYKIKIVERENLSY 273
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 350 CILDFCFEBSQPCVLSR-----SLQTTFLVNNKVFGLTHLMQDMVKDALRSFVDPV 404
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 274 CVLGG-----QRPMHLPMNQLASLGKTNBQSPHSQIHSTPIRNQV-PALQIPMSPL 325
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
QY 405 LSPK-----CYLYNN-----HOAKDCID-SFVTH-----CVR--PFCSL 435
Db 326 LSPKSPQVLRQOIAMAHLINQIIVASSELLAHQHPQAINQOFLNHPPIPRAVKPEPTNSS 385
QY 436 IQIHGHNRAQRQKLGHIIEFATLQDEFMTFFYNRAEKVDAALHTMLLKQEPORQHAC 495
Db 386 VEVSPIYQQVRDE---LKRASVSQAVFARVAFNRTO---GLLSEILRKEEDPR--- 433
QY 496 LGTWLVHNRIRIMQYLLSGFELE---LY-----SMHEYYIYWYLSLSE----- 535
Db 434 TASQSLLVNLRAM-QNFINLPEVERDRIYQDERERSMNPVNVSSASSPSSSRTTPOAK 492
QY 536 -----FLYAWLMSTLSRADGSMABEERIMEQOQGRSSKTKKK 574
Db 493 TSPTTDLPIKVDGANINITAAIYDEIQOEMKRAKVSQALFAKVAANKSQGLWCELLRWK 552
QY 575 KVRPLSREITMSQAYQNMCAKFKTMVAFMDGKVRKPKFELD---SEQVRYEHRFAPF 631
Db 553 ENPSPENRTIL---WENLC-----TIRRF-----LNLQHERDVIYEESRHHHSERMQ 597
QY 632 NSVMTTPP-PVHYLQ-----FKEMSDLNKYSPPQSPPELYVAASKHFQQAAMILENIP-- 682
Db 598 HVQLPPEPVQLHRQOSQAKSSPPREAPPPPTEDSCAKKPRSRKISLEALGIL 657
QY 683 -----NPDHEVNRIL--KVAKPNFVVMK 703
Db 658 QSFTHDVGLYPDQEAHITLSAQLDLPKHTTIK 689
```

RESULT 9

```
US-10-408-765A-696
; Sequence 696, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 696
; LENGTH: 761
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-696
```

```
Query Match 3.3%; Score 127; DB 16; Length 761;
Best Local Similarity 20.9%; Pred. No. 0.014;
Matches 132; Conservative 91; Mismatches 227; Indels 182; Gaps 31;

QY 217 BERPVEVL---RHQOCLAVFSRVKFTVLLTVLIAFTTKKTSVAEAKLMM----- 265
Db 95 EQLDGSLEYNDREHAEFVLVRKDVLSQVETALLALGYSHSSA-AQAQCIKLGKRWNP 153
QY 266 -----VOADLLSAIHSNLSHGIGIAQNDTTKGDHP-----IMMGPEPL--- 303
Db 154 LPLSVYTDAPDATVADMLOQVYHVVTLLKIQLSCKLEDLPACQWNHATVARNALKELLE 213
QY 304 VNOQLIPPTTPRYAKIKREEMNYFA-----RLIDRIKTV-CEVNVNLTNL--H 349
Db 214 MNQSTLAKECPLSQSMISSIVNSTYYANVSATKQCEFGRWYKTKIKVERVERENLSDY 273
QY 350 CILDDFFCFEBSQPCVLSR-----SLQTTFLVNDKNKVFTHLMQDMVKDALRFVDPVP 404
Db 274 CVLG-----QRPWHLPMNMQLASLTKTNEQSPHSQIHHSSTPIRNVQ-PALQIPMSFGL 325
```

```
QY 405 LSPK-----CYLYNN-----HOAKDCID-SFVTH-----CVR--PFCSL 435
Db 326 LSPKSPQVLRQOIAMAHLINQIIVASRLLAHQHPQAINQOFLNHPPIPRAVKPEPTNSS 385
QY 436 IQIHGHNRAQRQKLGHIIEFATLQDEFMTFFYNRAEKVDAALHTMLLKQEPORQHAC 495
Db 386 VEVSPIYQQVRDE---LKRASVSQAVFARVAFNRTO---GLLSEILRKEEDPR--- 433
QY 496 LGTWLVHNRIRIMQYLLSGFELE---LY-----SMHEYYIYWYLSLSE----- 535
Db 434 TASQSLLVNLRAM-QNFINLPEVERDRIYQDERERSMNPVNVSSASSPSSSRTTPOAK 492
QY 536 -----FLYAWLMSTLSRADGSMABEERIMEQOQGRSSKTKKK 574
Db 493 TSPTTDLPIKVDGANINITAAIYDEIQOEMKRAKVSQALFAKVAANKSQGLWCELLRWK 552
QY 575 KVRPLSREITMSQAYQNMCAKFKTMVAFMDGKVRKPKFELD---SEQVRYEHRFAPF 631
Db 553 ENPSPENRTIL---WENLC-----TIRRF-----LNLQHERDVIYEESRHHHSERMQ 597
QY 632 NSVMTTPP-PVHYLQ-----FKEMSDLNKYSPPQSPPELYVAASKHFQQAAMILENIP-- 682
Db 598 HVQLPPEPVQLHRQOSQAKSSPPREAPPPPTEDSCAKKPRSRKISLEALGIL 657
QY 683 -----NPDHEVNRIL--KVAKPNFVVMK 703
Db 658 QSFTHDVGLYPDQEAHITLSAQLDLPKHTTIK 689
```

RESULT 10

```
US-10-335-977-7409
; Sequence 7409, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragoras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 7409:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 969 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
```



```
;
; ORGANISM: Helicobacter pylori
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...969
; SEQUENCE DESCRIPTION: SEQ ID NO: 7409:
US-10-335-977-7409

Query Match          3.3%; Score 127; DB 12; Length 969;
Best Local Similarity 17.3%; Pred. No. 0.021;
Matches 143; Conservative 125; Mismatches 271; Indels 286; Gaps 37;

Qy 35 DFEACRELKLGELHDKLFGLEAMSALMMDPKMDAGMIGNOV-----NRK 82
Db 9 DYQECR-----DQILGVFK--GIDLREPENDIQRANPVFEIGAIKDLLENIE 56
Qy 83 VINFEQAIDGTIKIKDITLPELIGIMDT---CFCCILITWLEHSLAQTVFTCLY 134
Db 57 NLRSKQKITQSGVINGQSLNCDI--LMETGTGKTCFCL-----ECVYALHK 100
Qy 135 -IHNPDFIEDPAMKAFALGILKICDIAREKYNKAAVFEEDFQSMYGFKMANSVTDLRV 193
Db 101 NYHLSKFIVLAPSAIKGLVLSKITEITRE-----FFKSEYSNTHL-- 140
Qy 194 TGMCLKOVDDMQRRVKSTRSQGERDPEVELEHQOCLAVFSRVKFTFVLLTV----- 246
Db 141 -----ESYEDVERFILAS-----NHKCCVLVMTFSAFNKEKNTINQSCLEN 181
Qy 247 --LIAFTKETSAAVAEAKLMV-----QAADLLSAIHNSLHHGIQANDTTK 291
Db 182 TNLFNGAKSYNQALASMRPIVIMDEPHRFLGDKTKKYLEQLNAL--ITLRFGATFKDDY-- 238
Qy 292 GDHPIMMGFEPLVNQRLPPTFPFYAKIIKREEMVNYFARLID----- 334
Db 239 --HNLIVALD---SKAFDCAVKSIASVSGESDEYFLELKEANKKQNEATINVTTLN 293
Qy 335 -----RIKTCVENVNLNLHCLIDFCFBSQSPCVLSRLIQTTFVLVDNKKVFGTHLMQ 389
Db 294 KIQSVKVKTHDNLGALTHISALEDDYIVEIKTEI--RFLNGFNLLDQKEPF--SHLLE 349
Qy 390 D-----MKDALRSFVD-----PPVLSKPCVLYNNHQAKCIDSFVTHCVRP-----FC 433
Db 350 GEQEVMLKEAIKSHFEREEGLFKKGIKALCMVFING-----VNSYLSENEKPAKLALLFE 404
Qy 434 SLIQIHGHNRRARQDKLGHILEEPATLQDEFMTFVFNRAE---KVDAAHLHTMLLKQEPQ 489
Db 405 KLYQ-----QKLEVLKX---PLDENYRAVLETKAILKLVHGGYFAKSKKESDE 451
Qy 490 RQHLACLGTWVLYHNLRIMIQVLSGPELELYSMHEYYIYVYLSSEFLYAW---LMSTL 545
Db 452 VQVIA-----LILKEKEKLLSPDSDL---RFIFSQWALQE---GWDNPNVMTIC 494
Qy 546 SRADGSQMAEERIMEEQKGRSSKTKKKKKVR----- 578
Db 495 KLAPESS-----SNITKLOQIGRGLAVNDKGERITKEHADFDVFNELVVIVPQVEGDFVG 550
Qy 579 PLSEITWSQAVQNMCAQ-----MFKTMVAFMDGKVRKPKFELDSQV 622
Db 551 AIQOEISEHSLIKQVFSABEELEKSGMWKGYGVLFETLEGLGFGKETTDDNFNKLTLNQ- 609
Qy 623 RYEHRFAPFNSVMTPPP-----VHYLQFKMSDLNKYSPPPQSPELYVAASKHFOQAK 675
Db 610 -----NEFLKEPELEKLDKDEYLDPEKLDKDFLK-----DRLIG---HFR--- 646
Qy 676 MILENIPNPDENVNRLKVAQNFVVMKLLAGCHKKESKVPPEFD 720
Db 647 -----VRNKNERKTEKIKINKENFKKFTLWAGLNHQARIAYID 686

RESULT 11
US-10-335-977-7410
; Sequence 7410, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
```

```
;
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
;
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy B.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 7410:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 972 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...972
; SEQUENCE DESCRIPTION: SEQ ID NO: 7410:
US-10-335-977-7410
```

```
Query Match          3.3%; Score 127; DB 12; Length 972;
Best Local Similarity 17.3%; Pred. No. 0.021;
Matches 143; Conservative 125; Mismatches 271; Indels 286; Gaps 37;

Qy 35 DFEACRELKLGELHDKLFGLEAMSALMMDPKMDAGMIGNOV-----NRK 82
Db 12 DYQECR-----DQILGVFK--GIDLREPENDIQRANPVFEIGAIKDLLENIE 59
Qy 83 VINFEQAIDGTIKIKDITLPELIGIMDT---CFCCILITWLEHSLAQTVFTCLY 134
Db 60 NLRSKQKITQSGVINGQSLNCDI--LMETGTGKTCFCL-----ECVYALHK 103
Qy 135 -IHNPDFIEDPAMKAFALGILKICDIAREKYNKAAVFEEDFQSMYGFKMANSVTDLRV 193
Db 104 NYHLSKFIVLAPSAIKGLVLSKITEITRE-----FFKSEYSNTHL-- 143
Qy 194 TGMCLKOVDDMQRRVKSTRSQGERDPEVELEHQOCLAVFSRVKFTFVLLTV----- 246
Db 144 -----ESYEDVERFILAS-----NHKCCVLVMTFSAFNKEKNTINQSCLEN 184
Qy 247 --LIAFTKETSAAVAEAKLMV-----QAADLLSAIHNSLHHGIQANDTTK 291
Db 185 TNLFNGAKSYNQALASMRPIVIMDEPHRFLGDKTKKYLEQLNAL--ITLRFGATFKDDY-- 241
Qy 292 GDHPIMMGFEPLVNQRLPPTFPFYAKIIKREEMVNYFARLID----- 334
Db 242 --HNLIVALD---SKAFDCAVKSIASVSGESDEYFLELKEANKKQNEATINVTTLN 296
```


Db 2762 SDEDRLLVVFNRLGLTMTSEFNTL---HMMYHEATACHVTGDLVELLSIFLSVLKSTRPY 2818
QY 215 QGEERDPEVELEHQCLAVFSRVKTRVLLTVLIAFTKE--TSABAQK--LMVQAADL 271
Db 2819 LQRDKVKQALQWQE-----RIEFAHKLTLTNSYSPPELNACIDVLKELVLLSPHDF 2872
QY 272 LSAI-----HNSLHHGTAQNDTTKGHPIMMG--FEPLVNORLL-----PP----- 311
Db 2873 LHTLVPELQHNHCTYH-----SNI PMSLGPYPCRENKILGKSNIRPRPEL 2922
QY 312 ---TFPRVAKIIRKEMNYFARLIDRIKTVCVVNLNLHLCILDFPCE--FSEQSPCVLS 367
Db 2923 NMCLLETFWVETSGKDDV--YDRMLLDYFFSVHQFIHL---CRVAINCEKFT----- 2971
QY 368 RSLLOTTLVDNKKV-----FGTHLMQD---MVXDALRSFVDPVLSLSP--KC----- 409
Db 2972 -TLVKLSVLVAYEGPLHLAFLPKLWTELCTQOSAMSKNCIKLLCEDPVPFAEYIKILMD 3030
QY 410 ---LYLNHQAKDCIDSFVTHCVRPFCSLIQIHGHRARQRDKLG-----HILEEFATLQD 462
Db 3031 BRTPFLANN-----IVYTFMTH-----FLKQVQSQVSEANCANLSTLTITNLISQYNLQS 3081
QY 463 EPMYTFYFNRAE--KVDAALH-----TMLLKQEPQORHACLGTVWLYHNURIMITQYLLS 514
Db 3082 DFS-----NRVETSKASASLNGDLRALALLSVHTPKQ----- 3114
QY 515 GPELELYSMHEYYIYVYLSFEFLYAWLMSLTSRADGSQMABERIMBEEQKGRSSKTKKK 574
Db 3115 -----LNPALITPLQELLSKCRTCIQORNSLQEQEAKER---KT added 3153
QY 575 KKVRLPSREITMSQAYQNN--CAGMEKTMVAFDMDGKVRKPKPFELDSEQVRYEHRFAPFN 632
Db 3154 EGATPIKRRVSSDEHITVDSCISDMKTETR-----EVLTPITSTDNE-----TRDS 3200
QY 633 SVMTPPPVHYLOFKEMSLINKYS-----PPQSPELYVAASKHFQ 673
Db 3201 SIIDPTEQDLPSPESSSVKYEVRMEVPSFSDMSNIRSQHAE 3244

RESULT 14
US-10-412-897-3
; Sequence 3, Application US/10412897
; Publication No. US20030220224A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES ENCODING THE HUMAN CITRON KINASE
; FILE REFERENCE: D0193 NP
; CURRENT APPLICATION NUMBER: US/10/412,897
; PRIOR FILING DATE: 2003-04-11
; PRIOR FILING DATE: U.S. 60/372,745
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 3
; LENGTH: 1441
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-412-897-3

Query Match 3.0%; Score 116.5; DB 15; Length 1441;
Best Local Similarity 19.0%; Pred. No. 0.4;
Matches 150; Conservative 132; Mismatches 316; Indels 193; Gaps 39;

QY 20 EKEKSNNTNWVDITQDF-----BEACRELKGLBLLH---DKLFGLEFMAAIEEMDPKMDA 72
Db 73 ERAXKASTATELLQNIQAKERAEREL---EKLHNRDSSSEGIKKLVEAELESEKHRE 129
QY 73 GMIGNQ-----VNRKVLNFEQAIK--DGTIKIKDLTLPELLIGIMDTFCCLITWLEGH-- 123
Db 130 AQVSAOHLEHLVKQEQHVEEKIKVLNDQIK--KDLADKESLENM-----MORHEE 178
QY 124 -----SLAQTVFTCLYIHNPDFIEDPKAMKAFALGILKICDIAREKVNKAAYFEED 174

Db 179 EAHEKGKILSQKAMINAM-----DSKIRSEQRIVELSE-ANKLAANSSUFTQEN 228
QY 175 PQSMTYTGPMANSVTDLR-----VTGMLKQVEDDMQR---RVKSTRSR--QGE 217
Db 229 MKA-----QBEIMSELQKQFYLETQAGKLUFAQNRLKEEQLEKISHQDHSKSRILLELE 282
QY 218 ERDPEVELEHQCLAVFSRVKTRVLLTVLIAFTKETSAAVAEQKLMVQAADLLSAIHN 277
Db 283 TRLREVSLEHEE-----OKLELKROLTELQSLQERESQLTA-----LQAA--RAALES 329
QY 278 SLHHG--IOAQNDTTKGDPHPIMMGFEPFLVNORLLPPTFPYAKIIRKEMNYFARLIDRI 336
Db 330 QLQAKTELETTAEABEIQ-----ALTAHRDEI---QRKFAL 366
QY 337 KTVCEVNV--LTNLHCILDFCEPSESPCVLSRSLLOTTLVDNKKVFGTHLMQDMVKD 394
Db 367 RNSCTVITDLBEQLNQLTEDNAELNNQN--FYLKOL-----DEAGANDEIVQ----- 413
QY 395 ALRSFVD--PPVLSPKCYLVNHHQAKDCIDSFVTHCVRPFCSLIQIHGH--NRARQ--- 446
Db 414 -LRSEVDHLRREITEREMOLTSQKQTMALKTTCTMLEEQVLDLALMDLELEKERQWEA 472
QY 447 -RDKLGHILEEFATLQDFMTFFYFNRAEKVDAALHTMLLKQEPQORHACLGTVWLYHNL 505
Db 473 WRSVLGDEKSFQECVRRELQRLM--DTEKQSFARADQIRITESQVVELA-----VKEHKA 525
QY 506 RIM--IQYLLSFELELYSMHEYYIYVYLSFEFLYAWLMSLTSRADGSQMABERIMBROOK 564
Db 526 EIALQALKEQKLAESLSKLNLDLEKKHAMLENARSLOQKLETERELKORLLEEQAQ 585
QY 565 GRSSKTKKKKKVRPLSREITMSQAYQNNC--AGMEKTMVAFDMDGKVRKPKPFELDSEOV 622
Db 586 LQQQMDLQKNHIFR-----LTQGLQALDRADLKTTERS--DLE-----IQLENIQV 630
QY 623 RYEHFAPFNSVMTPPP--VHYLOFKEMSLINKYSPPQSPELYVAASKH----- 670
Db 631 LYSHEKVKMEGTISQOTKLIDFLQAKMDQPAKKKKVPLQYNELKLALEKEKARCAELEBA 690
QY 671 FQQAAMILENI-----PNPDHEVNRILKVAKNFV-----VKKLIA--GG 708
Db 691 LQKTRIELRSAREEAHRKATDHPHPSTPATARQOIAMSIVRSPEHQPFSAMSLAPPSS 750
QY 709 HKKESKVPPEF 719
Db 751 RKESSTPEEP 761

RESULT 15
US-09-964-956-40
; Sequence 40, Application US/09964956
; Publication No. US20040043926A1
; GENERAL INFORMATION:
; APPLICANT: Gerlach, Valerie L
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glenda
; APPLICANT: Millet, Isabelle
; APPLICANT: Stone, David
; APPLICANT: Gunther, Erik
; APPLICANT: Ellerman, Karen
; APPLICANT: Grosse, William M
; APPLICANT: Alsobrook II, John P
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Leach, Martin D
; APPLICANT: Shinkets, Richard A
; TITLE OF INVENTION: No. US20040043926A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-124
; CURRENT APPLICATION NUMBER: US/09/964,956
; CURRENT FILING DATE: 2001-09-26

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 21:50:17 ; Search time 19 Seconds
(without alignments)
1991.676 Million cell updates/sec

Title: US-10-001-857-145
Perfect score: 3849
Sequence: 1 MWKASVDDDDSGWELSMPE.....KVPPEFDSAHKYPVVKLV 733

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pgp:*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pgp:*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pgp:*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pgp:*
5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pgp:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pgp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	127	3.3	1089	1	US-08-375-300-2
2	127	3.3	1089	3	US-09-177-431-2
3	127	3.3	1089	5	PCT-US95-16930-2
4	113.5	2.9	1279	4	US-09-724-517-2
5	113.5	2.9	1279	4	US-09-641-807A-2
6	113.5	2.9	1279	4	US-09-723-096-2
7	112	2.9	1786	3	US-08-973-462-8
8	111.5	2.9	1332	2	US-08-971-244-2
9	111.5	2.9	1332	3	US-09-286-891-2
10	109.5	2.8	1066	3	US-09-541-782-8
11	109.5	2.8	1066	4	US-09-723-820-8
12	109.5	2.8	1066	4	US-10-270-085-8
13	108	2.8	1388	2	US-08-685-576-1
14	105.5	2.7	1507	4	US-09-914-259-37
15	105	2.7	526	4	US-09-442-100-12
16	105	2.7	526	4	US-08-939-106-12
17	105	2.7	526	4	US-09-442-102-12
18	104.5	2.7	585	4	US-09-107-532A-5494
19	104.5	2.7	1527	2	US-09-376-330-2
20	102	2.7	732	2	US-08-533-669A-18
21	102	2.7	732	4	US-09-307-143-4
22	102	2.7	732	4	US-09-183-861-18
23	102	2.7	732	4	US-09-022-763-18
24	102	2.7	732	4	US-09-551-974A-18
25	102	2.7	732	4	US-09-565-501A-18
26	102	2.7	732	4	US-09-639-206A-18
27	102	2.7	732	4	US-09-874-923-18

28	101.5	2.6	1248	2	US-09-080-897-2	Sequence 2, Appli
29	101.5	2.6	1248	3	US-09-323-735-2	Sequence 2, Appli
30	101.5	2.6	1315	3	US-08-899-595-3	Sequence 3, Appli
31	101.5	2.6	2471	3	US-09-112-450-4	Sequence 4, Appli
32	101.5	2.6	2471	4	US-09-419-291A-4	Sequence 4, Appli
33	101.5	2.6	2662	4	US-09-595-684B-31	Sequence 31, Appli
34	101.5	2.6	3878	4	US-09-914-259-11	Sequence 11, Appli
35	100.5	2.6	2165	1	US-08-514-975B-2	Sequence 2, Appli
36	100.5	2.6	2165	5	PCT-US95-12507-2	Sequence 2, Appli
37	99	2.6	1581	4	US-09-866-108A-15754	Sequence 15754, A
38	98.5	2.6	540	4	US-09-588-995A-115	Sequence 115, App
39	98.5	2.6	4536	4	US-09-180-422B-27	Sequence 27, Appl
40	98.5	2.6	4563	4	US-09-108-006C-1	Sequence 1, Appli
41	98	2.5	1306	3	US-08-999-774A-13	Sequence 13, Appl
42	98	2.5	3038	1	US-08-450-332-2	Sequence 2, Appli
43	98	2.5	3038	2	US-08-637-640-2	Sequence 2, Appli
44	98	2.5	3038	3	US-09-004-406C-2	Sequence 2, Appli
45	97.5	2.5	2482	1	US-08-328-254-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-375-300-2
; Sequence 2, Application US/08375300
; Patent No. 5679566
; GENERAL INFORMATION:
; APPLICANT: Feng, He
; APPLICANT: Jacobson, Allan S.
; TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE PRODUCTION IN
; TITLE OF INVENTION: THE ABSENCE OF NONSENSE-MEDIATED mRNA DECAY FUNCTION
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street Suite 3100
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/375,300
; FILING DATE: 20-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. P.
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 04020/046001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)542-5070
; TELEFAX: (617)542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1089 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-375-300-2

Query Match 3.3%; Score 127; DB 1; Length 1089;
Best Local Similarity 17.6%; Pred. No. 0.0033;
Matches 117; Conservative 124; Mismatches 202; Indels 222; Gaps 34;
QY 24 KNTNWV-----DITQFEACRELKLGELL-----HDKLGLGL 57
DB 35 KNTGFIKKKGFKVKGSESLKDLSEASLEKYLSEIIVTTECLINLVNKNDDVIAV 94

```
QY 58 EAMSAI-EMMDPKMDAGMIGNQVNRKVLNFEQAIKDGTIKIKDLTLPGLIGIMDTFCCL 116
Db 95 EIIISLHQRFNGRFTSPLLG-----AFLQAFENPSVDIES-ERDEL-----QR 136
QY 117 ITWLEGHSLAQVFTCLYIHN-----PDIE-----DPMKAFALGIK- 155
Db 137 ITRVGN--LRVFTLYLVGVFTLDDIESKDAIPNFKQTKGRKDLFLSILREILNY 193
QY 156 -----ICDIAREKVNK-AAVFEED--FQSMYGFKMANSVTDL-----RVTGM 196
Db 194 KFKLGFTTTIATAFIKKFAPLFRDDNSWDLIYDSKLKALQSLFKNFIDATPARATEL 253
QY 197 LKQVEDDMQRRVKSTRSQGEERDEVELEHQCLAVFSRVKFTVLLTVLIAFTKETS 256
Db 254 HKKV-NKLQRHQKQIRTKGRDBEYVE-EYDKLLPIFIRFK-----TS 295
QY 257 AVAAQKLMVQAADLLSAIHNSLHHGICQAQNDTTKGDHPIMMGPEPLVQRLPPTFFRY 316
Db 296 AITLGEFFKLEIPELEGASNDL-----KETA-----SPMITNQILPEN--- 334
QY 317 AKIKREEMVNYFARLIDRIKTCEV-----VNLTLNLCILDFFCFSEQSPCVL 366
Db 335 QRLWENEDTRKFYEILPDISKTVEBSQSSKTEKDSNVNSKNINL---FFTDL-EWADC-- 388
QY 367 SRSLLQTTFLVD--NKKVFGTHLMQDMVKD-ALRSFVDP-----PVLSPKCYLYNNHQA 417
Db 389 -----KDIIDDLNRYWSSYLDNKATRNILKFFMETQDWSKLPPVYS-RFIATNSKYM 440
QY 418 KDCIDSVFTHCVRPFCSLIQIHGHNRRARQDKLGHILEEFATLQDEFMTFFYFNRAEKYDA 477
Db 441 PEIVSEFINYLDNGFRS--QLHS-----NKINVKNI 469
QY 478 ALHTMLKQEPQROHLACLTGWLYHNLRIMIQYLLSGFELEYLSMHEYYIYWLSEFL 537
Db 470 IFFSEMIK-----FQLIPSFMIHFKIRILIMYQVNNVEILTV-----LLEHSGKFL 517
QY 538 YAWLMTLSRADGSMABERIMEEQGRSKTKKKKKVRLSREITMSQAYQNMCMAGM 597
Db 518 -----LNKPEYKELMEKV-----QLINKKNDKRLN--NMKSALENIITLL 558
QY 598 FKTWV 602
Db 559 YPPSV 563
```

RESULT 2

```
US-09-177-431-2
; Sequence 2, Application US/09177431
; Patent No. 6071700
; GENERAL INFORMATION:
; APPLICANT: He, Feng
; APPLICANT: Jacobson, Allan S.
; TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE PRODUCTION IN THE
; TITLE OF INVENTION: ABSENCE OF NONSENSE-MEDIATED mRNA DECAY FUNCTION
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09177,431
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/955,472
; FILING DATE:
```

RESULT 3

```
PCT-US95-16930-2
; Sequence 2, Application PC/TUS9516930
; GENERAL INFORMATION:
```

```
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/050001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-9806
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1089 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-177-431-2
```

```
Query Match 3.3%; Score 127; DB 3; Length 1089;
Best Local Similarity 17.6%; Pred. No. 0.0033;
Matches 117; Conservative 124; Mismatches 202; Indels 222; Gaps 34;
```

```
QY 24 KSNTNWV-----DITODFEBAACRELKIGELL-----HDKLFGLF 57
Db 35 KRNTGFIKLLKXKGVKSESSLLKDLSEASLEKYLSEIIVTTECLNLVKNKNDVIAAV 94
QY 58 EAMSAI-EMMDPKMDAGMIGNQVNRKVLNFEQAIKDGTIKIKDLTLPGLIGIMDTFCCL 116
Db 95 EIIISLHQRFNGRFTSPLLG-----AFLQAFENPSVDIES-ERDEL-----QR 136
QY 117 ITWLEGHSLAQVFTCLYIHN-----PDIE-----DPMKAFALGIK- 155
Db 137 ITRVGN--LRVFTLYLVGVFTLDDIESKDAIPNFKQTKGRKDLFLSILREILNY 193
QY 156 -----ICDIAREKVNK-AAVFEED--FQSMYGFKMANSVTDL-----RVTGM 196
Db 194 KFKLGFTTTIATAFIKKFAPLFRDDNSWDLIYDSKLKALQSLFKNFIDATPARATEL 253
QY 197 LKQVEDDMQRRVKSTRSQGEERDEVELEHQCLAVFSRVKFTVLLTVLIAFTKETS 256
Db 254 HKKV-NKLQRHQKQIRTKGRDBEYVE-EYDKLLPIFIRFK-----TS 295
QY 257 AVAAQKLMVQAADLLSAIHNSLHHGICQAQNDTTKGDHPIMMGPEPLVQRLPPTFFRY 316
Db 296 AITLGEFFKLEIPELEGASNDL-----KETA-----SPMITNQILPEN--- 334
QY 317 AKIKREEMVNYFARLIDRIKTCEV-----VNLTLNLCILDFFCFSEQSPCVL 366
Db 335 QRLWENEDTRKFYEILPDISKTVEBSQSSKTEKDSNVNSKNINL---FFTDL-EWADC-- 388
QY 367 SRSLLQTTFLVD--NKKVFGTHLMQDMVKD-ALRSFVDP-----PVLSPKCYLYNNHQA 417
Db 389 -----KDIIDDLNRYWSSYLDNKATRNILKFFMETQDWSKLPPVYS-RFIATNSKYM 440
QY 418 KDCIDSVFTHCVRPFCSLIQIHGHNRRARQDKLGHILEEFATLQDEFMTFFYFNRAEKYDA 477
Db 441 PEIVSEFINYLDNGFRS--QLHS-----NKINVKNI 469
QY 478 ALHTMLKQEPQROHLACLTGWLYHNLRIMIQYLLSGFELEYLSMHEYYIYWLSEFL 537
Db 470 IFFSEMIK-----FQLIPSFMIHFKIRILIMYQVNNVEILTV-----LLEHSGKFL 517
QY 538 YAWLMTLSRADGSMABERIMEEQGRSKTKKKKKVRLSREITMSQAYQNMCMAGM 597
Db 518 -----LNKPEYKELMEKV-----QLINKKNDKRLN--NMKSALENIITLL 558
QY 598 FKTWV 602
Db 559 YPPSV 563
```

APPLICANT: UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL
 TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE
 TITLE OF INVENTION: PRODUCTION IN THE ABSENCE OF
 TITLE OF INVENTION: NONSENSE-MEDIATED mRNA DECAY
 TITLE OF INVENTION: FUNCTION
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson
 STREET: 225 Franklin Street Suite 3100
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02110-2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/16930
 FILING DATE: 27-DEC-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/375,300
 FILING DATE: 20-JAN-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Fasse, J. Peter
 REGISTRATION NUMBER: 32,983
 REFERENCE/DOCKET NUMBER: 04020/046W01
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)542-5070
 TELEFAX: (617)542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1089 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US95-16930-2

Query Match	3.3%;	Score 127;	DB 5;	Length 1089;
Best Local Similarity	17.6%;	Fred.No. 0.0033;		
Matches	117;	Conservative 124;	Mismatches 202;	Indels 222; Gaps 34;
QY	24	KSNTNW----	---DITODFEACRELKIGELL---	---HDKLGFLF 57
Db	35	KRNITGFIKKLKGFEVKQSESLKDLSEASLEKVLSELIIVTVECLNVLNKNDVDIAAV	94	
QY	58	EAMSAI-EMMDPKMDAGMIGNQVNRKVLNPEQAIKGTIKIKIDTLPELIGIMDTCFCLL	116	
Db	95	EITISGLHQRENGRFTSPLLG----	---AFLQAFENFSVDIES-ERDEL---	---QR 136
QY	117	ITWLEHSLAQVTFTCLYIHN----	---PQFIE-----DPAMKAFALGILK-	155
Db	137	ITRVKGN---LRVFTELYLVGVFTLDDISKDAIPNLFQKKTGRKDPFLFSILREILNY	193	
QY	156	-----ICDIAREKVNK-AAVFEED-----FOSMTYGFKMWANSVTDL-----	RVTGM 196	
Db	194	KFKLGTFTTTIATAFIKFPAPLFRDDNSWDLLIYDSKLKALQSLFKNFIDATPARATEL	253	
QY	197	LKDVEDDMQRVKSTRSQGEERDPFEVELEHQOCLAVFSRVKFTFVLLTVLIATTKETS	256	
Db	254	HKKY-NKLQREHQRCQITGKLRDYVE- EYDKLLPIFIRFK-----	TS 295	
QY	257	AVABEAQKLMVQAAQLLSAIHNSLHHGIQAQNDTTKGDHPIMMGPEPLVNQRLPPTFPFRY	316	
Db	296	AITLGEFFKLEIPELEGASNDL-----	---KETA-----SPMITNQILPPN----	334
QY	317	AKILIREBMVNYFARLIDRIKTVCEV-----	---VNLTNLHCILDFCFSESQSPCVL 366	
Db	335	ORLWENEDTRAFYEILPDIISKTVEESQSKTEKDSNVNSKNIL---	---FPTDL-EMADC-- 388	

[illegible][illegible]

```

Db 773 PKAEDLDACNLKRRKSGFSIDH-LQKLDQKQKWLDEEV-----EKLVLNQRQLELEADL 827
QY 313 PPRYAKIIKREBMVNYFARLIDRIKTVCEVNLTLHC-----ILDFCFESEQSPCVLS 367
Db 828 KKREAIIVSKKALLQEKSHLENKLRSSQALNTDSIKISTRNLNLE--QELSEKN-----880
QY 368 RSLQTTFLVDNKKVFGTHLMDVMKDAL-----RSFVDPDP-----VLSPK-----CYLYNN 414
Db 881 -VOLQTSAAEKTIKISEQVEVLQKEKQOLQRRHRVDKLNKGRVLSPEEHLVQLDEEG 939
QY 415 HOAKDCIDSFVTHCVRPFCSLIQIHGHNARQRDKLGHILEEFATLQD-EFMTF---YFN 470
Db 940 IEALEAAIEYRNESIQNRQKSLRASFNLSRGE---ANVLEKACLSPVEIRTLFRYFN 996
QY 471 RAEKVDAALHTMLLKQEPQOHL-----ACLGTVLYHNLIRIMIQVLLSGFELE 519
Db 997 KVMNLREARQQLYNEEMKMKVLERDNMVRLESALDHLKLCQDRRLTLQOKEHEQKMQ 1056
QY 520 LYSMEHYVYIYWLSEFLYAWLMSTL-SRADGSMABERIMEEQOQGRSSKTKKKKKVR 578
Db 1057 LLLHH-----FKEQDGEIMETFKTYEDIQOLEKDLFYFKKTSRDHKKKLKELVGE 1108
QY 579 PLSREITMSQAYQNNCAGMFKT---MVAFDMDGKVRKPKF-----ELDSE-----620
Db 1109 AIRQLAPSE-YQEAGDGLVPEGGGMLSEBKWASRPESMKLSGRBREMDSASSLRTO 1167
QY 621 ---QVRVEH--RFAPFNSVMTPPVHYL--QFKEMSDLNKYS 655
Db 1168 PNPQKLWEDIPELPPIHSSLAPPSGHMLGNENKTETDDNQFT 1209

RESULT 5
US-09-641-807A-2
; Sequence 2, Application US/09641807A
; Patent No. 6440731
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: No. 6440731el motor proteins and methods for
; FILE REFERENCE: 1031
; CURRENT APPLICATION NUMBER: US/09/641,807A
; CURRENT FILING DATE: 2000-08-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1279
; TYPE: PRT
; ORGANISM: Human
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (409)...(446)
; OTHER INFORMATION: Xaa = any amino acid
US-09-641-807A-2

Query Match 2.9%; Score 113.5; DB 4; Length 1279;
Best Local Similarity 20.3%; Pred. No. 0.092;
Matches 155; Conservative 125; Mismatches 291; Indels 191; Gaps 39;

QY 8 DDDSGWELSMPEKMEKXNT-----NWVDITQDFEEACRELKGLGELLHDKLFGLEFAMS 62
Db 525 DDEES-----EQEKSQGTCSRWSI---QKPDVSVCISVELSDTQDE-----TQKSD 568
QY 63 IEMDPKMDAGMIGNQV--RKVLNFEQAIGTKIKDLT-----LPFLIGI-----108
Db 569 LENEDLKIDCQESQELNQLKXSERILTEAKQWRELINIKWKEDLIKELIKTGND 628
QY 109 --MDTCCLLTWLEGS-----LACTVFTCLYHNPFDIEDPAMKAFGLIKIDIA 160
Db 629 KSVSKQSLVKLEHDAEQAKVELIETQKQLEKNK-LSDVAMK-----VKLOKEF 681
QY 161 REKNVKAAPFEEDFQSMYTGFKMANSVTDLRVTGMLKDVEDDDQRRVKSTRSQGEERD 220

```

```

Db 682 RKKMDAA-----KLFVQVLQKQKQDS-----KKLASLSIQNEKRA 716
QY 221 PEVE--LEHQOCLA VFSRVKTRVL-----LTVLIAFTKKETS AVAAQ-----KLM 265
Db 717 NELEQSVDHMK-----YQIKIQRLKRENEKRRKQDAVIKRDQKQIKETQLTKTQEEGLK 772
QY 266 VQAADL-----LSAIHNSLHHGIAQNDTTKG--DHPIMMGFEPLVNOR-----LLPPT 312
Db 773 PKAEDLDACNLKRRKSGFSIDH-LQKLDQKQKWLDEEV-----EKLVLNQRQLELEADL 827
QY 313 PPRYAKIIKREBMVNYFARLIDRIKTVCEVNLTLHC-----ILDFCFESEQSPCVLS 367
Db 828 KKREAIIVSKKALLQEKSHLENKLRSSQALNTDSIKISTRNLNLE--QELSEKN-----880
QY 368 RSLQTTFLVDNKKVFGTHLMDVMKDAL-----RSFVDPDP-----VLSPK-----CYLYNN 414
Db 881 -VOLQTSAAEKTIKISEQVEVLQKEKQOLQRRHRVDKLNKGRVLSPEEHLVQLDEEG 939
QY 415 HOAKDCIDSFVTHCVRPFCSLIQIHGHNARQRDKLGHILEEFATLQD-EFMTF---YFN 470
Db 940 IEALEAAIEYRNESIQNRQKSLRASFNLSRGE---ANVLEKACLSPVEIRTLFRYFN 996
QY 471 RAEKVDAALHTMLLKQEPQOHL-----ACLGTVLYHNLIRIMIQVLLSGFELE 519
Db 997 KVMNLREARQQLYNEEMKMKVLERDNMVRLESALDHLKLCQDRRLTLQOKEHEQKMQ 1056
QY 520 LYSMEHYVYIYWLSEFLYAWLMSTL-SRADGSMABERIMEEQOQGRSSKTKKKKKVR 578
Db 1057 LLLHH-----FKEQDGEIMETFKTYEDIQOLEKDLFYFKKTSRDHKKKLKELVGE 1108
QY 579 PLSREITMSQAYQNNCAGMFKT---MVAFDMDGKVRKPKF-----ELDSE-----620
Db 1109 AIRQLAPSE-YQEAGDGLVPEGGGMLSEBKWASRPESMKLSGRBREMDSASSLRTO 1167
QY 621 ---QVRVEH--RFAPFNSVMTPPVHYL--QFKEMSDLNKYS 655
Db 1168 PNPQKLWEDIPELPPIHSSLAPPSGHMLGNENKTETDDNQFT 1209

RESULT 6
US-09-723-096-2
; Sequence 2, Application US/09723096
; Patent No. 6448026
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: No. 6448026el motor proteins and methods for
; FILE REFERENCE: 1031
; CURRENT APPLICATION NUMBER: US/09/723,096
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US/09/641,807
; PRIOR FILING DATE: 2000-08-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1279
; TYPE: PRT
; ORGANISM: Human
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (409)...(436)
; OTHER INFORMATION: Xaa = any amino acid
US-09-723-096-2

Query Match 2.9%; Score 113.5; DB 4; Length 1279;
Best Local Similarity 20.3%; Pred. No. 0.092;
Matches 155; Conservative 125; Mismatches 291; Indels 191; Gaps 39;

QY 8 DDDSGWELSMPEKMEKXNT-----NWVDITQDFEEACRELKGLGELLHDKLFGLEFAMS 62
Db 525 DDEES-----EQEKSQGTCSRWSI---QKPDVSVCISVELSDTQDE-----TQKSD 568

```


QY	63	IEMMDPKMDAGMIGNOVN-RKVUNFFQALIKDGTIKIKILT-----LPBLIGI----	108
Db	569	LENEDLIKIDLOESQBELINQKLNSEIRIITEAKQKMRLEITINIKMKEDILIKELITGNDA	628
QY	109	--MDTFCCLITWLEGHS-----LAQTVFTCLYIHNPDFIEDPAMKAFALGIKICDIA	160
Db	629	KSVSKOYSLKVTKLEHDACAKVELLETQKOLENKO-LSDVANK-----VKLOKEF	681
QY	161	REKWNKAAPVEEBDFOSMTYGFKAANSVTDLRVTGMKDOVEDDMQRRVKSTRSQEERD	220
Db	682	RKMDAA-----KLRQVQLQKXQODS---KKLASISIQNEKRA	716
QY	221	PEVE--LEHOOCCLAVFSRVKFTVL-----LTVLIATFKETSAVAEAQ-----KLM	265
Db	717	NELEQSVDMK-----YQKIQLOQRKLBENEKRQOLDAVIKRQOKIKEIQLTKGQEGELK	772
QY	266	VQAADL-----LSAIHNSLHHGICQAONDTTKG-DHPIMMGFEPIJVNOR-----LLPPT	312
Db	773	PKAEEDLDACNLKRKRGSGFSIDH-LQKLEQKWLDEEV-----EKVINORQELELEADL	827
QY	313	FPYAKIIEBENMVNFARLIDRIKTVCEVWNLTNLHC-----ILLDFCFEFSQSPCVLS	367
Db	828	KKREAVISKKBALQSKSHLENKURSSQAINTDSUKISTRMLLE--OELSEKN-----	880
QY	368	RSLLQTTFLVDNKKVFGTHLMQDMVRDAL---RSFVDPDP-----VLSPK-----CYLYNN	414
Db	881	-VOLQTSABEKTIKSEQVEVLQKEQDQLOKRRHVDDEKLKNGRVLSPEEHVLFQLEEG	939
QY	415	HOAKDCIDSFVTHCVPPFCSLIQIHGHNARQDKLGHILEEATLQD-EFMTF---YFN	470
Db	940	IEALEAIEYRNEISIQNRQKSLRASHFNSRGE---ANVLEKLUACLSPVIRITILFRYFN	996
QY	471	RABKVDAALHTMLLKQBPQOHL-----ACLGTVLVYHNLRIMIQIYLLSGFELE	519
Db	997	KVYNLREAFERKQQLYNEEMKMKVLERDNMVRELESALDHLKLCQDRRLTLQKEHEBQKQ	1056
QY	520	LYSMHEYYYIYWLSEFLYAWLMSLT--SRADGSGMAEERIMEQOQGRSKTKTKKKKVR	578
Db	1057	LLLPH-----FKEQDGEGIMETFKTYEDKIQOLEKOLIFYKTSRDKHKKKLKELVGE	1108
QY	579	PLSREITMSQAYNMCMGFMKT---MVAFDMQGVKVRKPF-----ELDSE-----	620
Db	1109	AIRQLAPSE-YQEAGDGVLPKPGGMLSEELKWA SRPESMKLSGRENEMDSASSIRTO	1167
QY	621	---QVRVEH--RFAPFNSVMTPPPVHYL---QPKEMSDLNKYS	655
Db	1168	PNPKLMWEDPELPPHSHSLAPSGHMLGNENKETDNDQFT	1209

RESULTS

```

US-08-973-462-8
; Sequence 8, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 8
; LENGTH: 1786
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-8

```

Query Match	2.9%;	Score 112;	DB 3;	Length 1786;
Best Local Similarity	18.1%;	Pred. No. 0.22;		
Matches 129;	Conservative 114;	Mismatches 227;	Indels 242;	Gaps 30;
QY	17	SMPEKMEKNTNWVDITQDFEACRELKLGELH	----	DKLF-----G 55
Db	966	NVAENLEKUNETVNTVLQKVETVEIS	-----GSELENEMDKAFSEIFDNVKGIOENLLTG	1024
QY	56	LFEAM--SAIEMDPKQDAGMIGNQNRKVLNFEQAIDGTI	----	KIKDNLPELIGIMDTC 112
Db	1025	MFRESIETSVIQSEKVD--LNEENVSSILDNIENKEGLLKLENISS	-----	TS--GVOET- 1078
QY	113	FCCLITWLESHLAQVTCFLYIHNPDFIEDPAMKAFALGILK	-----	----- 155
Db	1079	-----VTEHVEQNYYVD--VDVPAMKQDFLIGLNEAGLKEMFENLSDVFK	1122	
QY	156	-----ICDIAREKYNK-----AAVFE-----	----	REDFQSMTYGFQMANSVTD-- 190
Db	1123	SESDVITVEIKEDPQKEVEKETVSIIEEMEENIVDVLEEKKEDUT	----	DKMDAIVAEESI 1180
QY	191	-----LRVTGMLKDVEDDMQRRVKSTRSQGGERDPEV	----	ELEHQQCLAVFSRKFTRVL 243
Db	1181	EISSDSKEETESIKKDKVSLVVEEQDNDMDSEVKEVLKKNMEELMKOAVEINDI	----	----- 1239
QY	244	LTVLIIFTKETSATAEAKIMVQAAADLSAIHNSLHHGIQANDT	-----	TKGDHPIMMG 299
Db	1240	TSKLIBETQELNEVEADLIKMDKLEKLEKALSSESKIIDA	-----	DKDDTLEKVIIEEHDITTT 1299
QY	300	FELVNRQLLPPFPYAKTIIEEEMVNYFARLIDRIKTVCEVVNLNLHCILDFPCEFS	359	
Db	1300	LDEVELKQVED-----KIEKVSDDKLUDEEDILKEVEKELKELESE	-----	LEEDKEL- 1348
QY	360	EQSPCVLSRLLOTTFLVDNKKVFGTHL-----	----	MQDMVKDALRSFVDPVPVLSPKCY 410
Db	1349	-----KTIEDTILKEKTEKIDHFEKFEFESEIKDEADILKEVSSLEVEEEK-K	1398	
QY	411	LYNNHQAKOCIDSFVTHCVRPFCSLIQIHGHNRAORDKI	-----	GHILE-----EF 457
Db	1399	LEEVELKEERVE-----HIISGDAHIKGLSEDDLEEDVDLKGSLDKMLKGEMEL	1447	
QY	458	ATLQDEPM-----TFYFNRAEKVDAALHTMLLKQBPQOHLACLGTVVLNHLRIMTQYLL	513	
Db	1448	GDMDKESLEDVTTKLGERVERESLKDVLSSALGMDBEQ	-----	----- 1483
QY	514	SGFELELYSMHEYYIYVWYLSFLYAWLMSLGRAGSQMAEERIMEEQKGRSSKKTKK	573	
Db	1484	-----WKTRKKAQRPKL-EVVLKEEVEEPKKKIITK	1514	
QY	574	KKKVRPLRSREITMSQAYQNNCAGMFKTMWAFMDGKVRKPKFELDSEQVRYE	625	
Db	1515	KK-----VFDTI--KKQKPKDSIVEVMKDE	1538	

RESULT 8

```

US-08-971-244-2
; Sequence 2, Application US/08971244
; Patent No. 5891719
; GENERAL INFORMATION:
; APPLICANT: Cohen, Lucy
; APPLICANT: Baerlele, Patrick
; TITLE OF INVENTION: IKAP Proteins, Nucleic Acids and Methods
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/971,244
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: T97-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1332 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-971-244-2

Query Match          2.9%; Score 111.5; DB 2; Length 1332;
Best Local Similarity 18.7%; Pred. No. 0.16;
Matches 145; Conservative 110; Mismatches 252; Indels 269; Gaps 40;

Qy 70 MDAGMIGNQVN-----RKVLNFEQAIDGTIKID-----LTLP-----ELIGIMTCFCCL 116
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
673 LQAGLSSNHVSHGVELRKV---ERGSRIVTVPQDTKLVLQMPRGNLEVVHRLVLAQI 729
Qy 117 ITWLEHSLAQTVTC-----LYIHNPDIEDPAMKAFALGILKICDIAREKVNK 166
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
730 RKWLD-KLMFKEAECMKRLINLPIYDHN-----KVF-LGNVETFIKQIDSVNH 779
Qy 167 AAVF-----EEDFOSMTYGFKMANSVTDLR-VTGMLKDVEDDMORRVKSTRSQEERDP 221
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
730 RKWLD-KLMFKEAECMKRLINLPIYDHN-----KVF-LGNVETFIKQIDSVNH 779
Qy 167 AAVF-----EEDFOSMTYGFKMANSVTDLR-VTGMLKDVEDDMORRVKSTRSQEERDP 221
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
780 INLFFTELKEEDVTKMYPAVTSVYLSRDPDGNKIDLVCDAMRAVM-----ESINP 832
Qy 222 EVELHQQCLAVF-SRVKFTVRLTLVLI-----AFTKKEISAVAEAKLMVQAADL 271
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
833 -----HKYCLSILTSHVKKTTPELEIVLQVHEIQGNAPSDPDVSAEALKYLHLVDV 887
Qy 272 LSAIHNSLHHGICQAQNTTKGDHIMMG-----PEPLVNO-RLLPPTFPFYA--KI 319
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
888 NELYDHSI-----GTYDFDLVLMVAEKSKQPKYLPFLNLTKKMETNYQFTIDKY 939
Qy 320 IKREEMV-----NYFARLIDRIKTVCEVNLTLNHLILDFFCFESQSPCVLSRSL 370
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
940 LKRYEKAIGHLSKCGPEYFPECLNLIKD-----KNLYNE--ALKLYSPSSQO----- 984
Qy 371 LQTFELVDNKKVFGTHLMQDMVKDALRSFVDPVLSPKYLYNNHQAQKDCIDSFVT--HC 428
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
985 -----YQDISIAYGEHLMQE-----HMYEPAGLMFARCGAH-----EKALSAFLTQGNW 1028
Qy 429 VRPFCSLIQIH-----GHNRA-----RQDKLGHILEEATLQDEF----- 464
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
1029 KQALCVAAQLNFTKQDLVGLGRTLAGKLVEGRKHIDAMVLEBCAQDYEEAVLLEGA 1088
Qy 465 -----MTFYFNRAEKVDAAHLHTMLLKQEPQORHACLIGTWVL---YHNLRIMTQYLLS- 514
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
1089 WEERALRVYKYNRLDIETNVKPSIL--EAQKNYMAFLDSQTATPSRHKKELLVVRELKE 1146
Qy 515 -----GFELELYSMHEYYIYVYLSEFLYAWLMSTLSRADGSGM-----AE 555
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
1147 QAOQAQLDDEVPHQESDLFS-----FTSSVVSQSGSEMSGKSHSN 1186
Qy 556 ERIMEEQKGRSSKTKTKKKK-----VRPLSREITWSQAYQNMCAGMFKTMV 602
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
1187 SRISARSKNR--RKAERKSHLKEGSPLEDLALIEALSEVVQNTENLKDVEYHILKVL 1244
Qy 603 AFQMD--GKVRKPKFELDSEQVRYEHRFAPFNSVMTPPPVHYLQFKEMSLNKYSPPPPQS 660
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
1245 LFEDFEGRELQKAFE-----DTLQLMERS-----L 1270
```

```

Qy 661 PELYVAASKHFQQAQKMLNIPNPDHEVNRILKVAKPNFVVMKLLAGGHKESKVP 716
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
1271 PEIWTLT---YQQ-----NSATPVLGPNSTANSINASYQQOKTSVP 1308

RESULT 9
US-09-286-891-2
; Sequence 2, Application US/09286891
; Patent No. 6172195
; GENERAL INFORMATION:
; APPLICANT: Cohen, Lucy
; APPLICANT: Baeuerle, Patrick
; TITLE OF INVENTION: IKAP Proteins, Nucleic Acids and Methods
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/286,891
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/971,244
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: T97-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1332 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-286-891-2

Query Match          2.9%; Score 111.5; DB 3; Length 1332;
Best Local Similarity 18.7%; Pred. No. 0.16;
Matches 145; Conservative 110; Mismatches 252; Indels 269; Gaps 40;

Qy 70 MDAGMIGNQVN-----RKVLNFEQAIDGTIKID-----LTLP-----ELIGIMTCFCCL 116
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
673 LQAGLSSNHVSHGVELRKV---ERGSRIVTVPQDTKLVLQMPRGNLEVVHRLVLAQI 729
Qy 117 ITWLEHSLAQTVTC-----LYIHNPDIEDPAMKAFALGILKICDIAREKVNK 166
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
730 RKWLD-KLMFKEAECMKRLINLPIYDHN-----KVF-LGNVETFIKQIDSVNH 779
Qy 167 AAVF-----EEDFOSMTYGFKMANSVTDLR-VTGMLKDVEDDMORRVKSTRSQEERDP 221
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
780 INLFFTELKEEDVTKMYPAVTSVYLSRDPDGNKIDLVCDAMRAVM-----ESINP 832
Qy 222 EVELHQQCLAVF-SRVKFTVRLTLVLI-----AFTKKEISAVAEAKLMVQAADL 271
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
833 -----HKYCLSILTSHVKKTTPELEIVLQVHEIQGNAPSDPDVSAEALKYLHLVDV 887
Qy 272 LSAIHNSLHHGICQAQNTTKGDHIMMG-----PEPLVNO-RLLPPTFPFYA--KI 319
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
888 NELYDHSI-----GTYDFDLVLMVAEKSKQPKYLPFLNLTKKMETNYQFTIDKY 939
Qy 320 IKREEMV-----NYFARLIDRIKTVCEVNLTLNHLILDFFCFESQSPCVLSRSL 370
```

```
Db 940 LKRYEKAIGHLSKGPYFFECNLIIKD-----KULYNE--ALKLYSPSSQ----- 984
Qy 371 LQTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVLPSPKCYL-----YNNHQAOKCIDSFT--HC 428
Db 985 -----YQDISIAYGEHLMQE-----HMYEPAGLWFARCGAH-----EKALSAFLTCGNW 1028
Qy 429 VRPFCSLIQIH-----CHNEPA-----RQDKLGHILEEFATLQDEF----- 464
Db 1029 KQALCVAQNLNFTKQIIVGLGRTLAGKLVEQRKHIDAMVLEECQAQDYEEAVLLLLSGAA 1088
Qy 465 -----MTFVFNRAEKYVDAALHTMLLKQEPOROHLAGICGTWVL---YNNLRIMIQYLLS- 514
Db 1089 WEEALRLVYKYNRLDIETNVKPSIL--EAQNTYMAFLDSQTATFSRHKKELLVVRLEKE 1146
Qy 515 -----GFELEYSMHBYIYIYVLSFLYAWLMSLTSLRADGSM-----AE 555
Db 1147 QAQOAGLDDEVPHQESDLFS-----ETSSVVSSEMSGKYSHN 1186
Qy 556 ERIMEEQKGRSSKTKKKK-----VRPLSREITMSQAYQNWCAGMFTMV 602
Db 1187 SRTSARSKNR--RKAERKXHSLEKGSPLDLALLEALSEVVQNTENLKDSVYHILKVLV 1244
Qy 603 AFQMD--GKVRKPKFELDSQVRYEHRFAPFNSVMTPPVHYLQFKXMSDLNKYSPPPQS 660
Db 1245 LFEFDEQRELOKAFE-----DTQLMERS-----L 1270
Qy 661 PELVAASKHPQQAAMILENIPNDPHEVNRILKVAKNFVVMKLLAGHKHESKVP 716
Db 1271 PEIWTLT---YQQ-----NSATPVLGPNSTANSIMASYQQOKTSVP 1308

RESULT 10
US-09-541-782-8
; Sequence 8, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/541,782
; CURRENT FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-541-782-8

Query Match 2.8%; Score 109.5; DB 3; Length 1066;
Best Local Similarity 19.7%; Pred. No. 0.17;
Matches 85; Conservative 80; Mismatches 168; Indels 99; Gaps 21;
Qy 122 GH-SLAQTVFTCLYIH-----NPDFIEDPAMKAFALGILKI-CDIAREKVNKAA 168
Db 336 GHKOIEETLSLEYAHRAKNIQNKPEVNQKLTKKTVLKEYTEEDIKLRDLMAARDKNGI 395
Qy 169 VFEEDFQSMTYGFKWANSVTDLRVGMLKVEDDMQRRVK--STRSRQGEERDPEVELE 226
Db 396 YLAETTYGEITLKLESQNRNELNEKML-LLKALKDELQNKKEKIFSEVMSLVKETOELKKT 454
Qy 227 HQQCLAVFSRVKFTVLLT-----VLIATFKKETSAAVAQAKLMVQAADLLSAI 275
Db 455 EENLNTKGTLLLTKKVLTKTKRYKKEKVELVASHMKTEQVLTTOAQEILA-AADLATDD 513
Qy 276 HNSLHHGTAQNDTTKGDPHIMMGFEPLV---NORLLPPTFPY-----AKIKREE 324
Db 514 THQLHGTIERREL---DEKIRSCDQFKDRMQDNLEMIGGSLNLYQDQQAALKEQLSQE 570
Qy 325 MVN-----YFARLIDRIKTVK--EVVNLTNLH-CILDFFCFEFSQSPCVLSRSL 371
Db 571 MVNSSVYSQRLALNSSKSIEMLEKEMCAQSLQDQTNLHNKLIQEVNWKISDQ----- 620
Qy 372 QTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVLPSPKCYL-----YNNHQAOKCIDS- 423
Db 621 -----HSQAFVAKLMEQMOQQOQ-----LMSKEIQTNLQVTEENNQRHKAMLDSMQ 666
Qy 424 --FVTHCVRPFCSLIQIHGHNRRARQDKLGH---LEEFATLQDEFMTFFYFN--ABKV 475
Db 667 EKPAT-IIDSLSQSVVEEAKQMKHKLQGLGAMSLPDAEELQNLQOELAN---ERALAQQE 722
Qy 476 DAALHTMLLKOE 487
Db 723 DALLESMWQME 734
```

```
Db 571 MVNSSVYSQRLALNSSKSIEMLEKEMCAQSLQDQTNLHNKLIQEVNWKISDQ----- 620
Qy 372 QTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVLPSPKCYL-----YNNHQAOKCIDS-- 423
Db 621 -----HSQAFVAKLMEQMOQQOQ-----LMSKEIQTNLQVTEENNQRHKAMLDSMQ 666
Qy 424 --FVTHCVRPFCSLIQIHGHNRRARQDKLGH---LEEFATLQDEFMTFFYFN--ABKV 475
Db 667 EKPAT-IIDSLSQSVVEEAKQMKHKLQGLGAMSLPDAEELQNLQOELAN---ERALAQQE 722
Qy 476 DAALHTMLLKOE 487
Db 723 DALLESMWQME 734

RESULT 11
US-09-723-820-8
; Sequence 8, Application US/09723820
; Patent No. 6468760
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/723,820
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-723-820-8

Query Match 2.8%; Score 109.5; DB 4; Length 1066;
Best Local Similarity 19.7%; Pred. No. 0.17;
Matches 85; Conservative 80; Mismatches 168; Indels 99; Gaps 21;
Qy 122 GH-SLAQTVFTCLYIH-----NPDFIEDPAMKAFALGILKI-CDIAREKVNKAA 168
Db 336 GHKOIEETLSLEYAHRAKNIQNKPEVNQKLTKKTVLKEYTEEDIKLRDLMAARDKNGI 395
Qy 169 VFEEDFQSMTYGFKWANSVTDLRVGMLKVEDDMQRRVK--STRSRQGEERDPEVELE 226
Db 396 YLAETTYGEITLKLESQNRNELNEKML-LLKALKDELQNKKEKIFSEVMSLVKETOELKKT 454
Qy 227 HQQCLAVFSRVKFTVLLT-----VLIATFKKETSAAVAQAKLMVQAADLLSAI 275
Db 455 EENLNTKGTLLLTKKVLTKTKRYKKEKVELVASHMKTEQVLTTOAQEILA-AADLATDD 513
Qy 276 HNSLHHGTAQNDTTKGDPHIMMGFEPLV---NORLLPPTFPY-----AKIKREE 324
Db 514 THQLHGTIERREL---DEKIRSCDQFKDRMQDNLEMIGGSLNLYQDQQAALKEQLSQE 570
Qy 325 MVN-----YFARLIDRIKTVK--EVVNLTNLH-CILDFFCFEFSQSPCVLSRSL 371
Db 571 MVNSSVYSQRLALNSSKSIEMLEKEMCAQSLQDQTNLHNKLIQEVNWKISDQ----- 620
Qy 372 QTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVLPSPKCYL-----YNNHQAOKCIDS- 423
Db 621 -----HSQAFVAKLMEQMOQQOQ-----LMSKEIQTNLQVTEENNQRHKAMLDSMQ 666
Qy 424 --FVTHCVRPFCSLIQIHGHNRRARQDKLGH---LEEFATLQDEFMTFFYFN--ABKV 475
Db 667 EKPAT-IIDSLSQSVVEEAKQMKHKLQGLGAMSLPDAEELQNLQOELAN---ERALAQQE 722
Qy 476 DAALHTMLLKOE 487
Db 723 DALLESMWQME 734
```

US001.13
US-08-685-576-1
Sequence 1, Application US/08685576
Patent No. 5906819
GENERAL INFORMATION:
APPLICANT: Kaibuchi, Koza
APPLICANT: Iwamatsu, Akihiro
APPLICANT: Nakano, Masaki
APPLICANT: Ito, Masaki
APPLICANT: Takahashi, No. 5906819

Db 903 -----ELQDER-----DSLAAQLEITLTKADSEQ 926

Qy 553 MAERIMEEQKGRSKTKKKKKVRPL-----SREITMSQA 589
 Db 927 LARS-IAFEQYSDLEKIMKEIKEMMARHKEQELTEKDA 966

RESULT 14

US-09-914-259-37
 ; Sequence 37, Application US/09914259
 ; Patent No. 6495336
 ; GENERAL INFORMATION:
 ; APPLICANT: Makowski, Lee
 ; APPLICANT: Hyman, Paul
 ; APPLICANT: Williams, Mark
 ; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
 ; FILE REFERENCE: 8471-010-999
 ; CURRENT APPLICATION NUMBER: US/09/914,259
 ; CURRENT FILING DATE: 2000-11-21
 ; NUMBER OF SEQ ID NOS: 180
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 37
 ; LENGTH: 1507
 ; TYPE: PRT
 ; ORGANISM: Drosophila melanogaster
 US-09-914-259-37

Query Match 2.7%; Score 105.5; DB 4; Length 1507;
 Best Local Similarity 17.2%; Pred. No. 0.74;
 Matches 107; Conservative 101; Mismatches 242; Indels 173; Gaps 25;

Qy 189 TDLEVTGMLDVEDDMQRRVKSTRSGEERDPEVELEHQCCLAVFSTRVRL----- 243
 Db 55 TNLTAGKPKKRRNNKEKESDAARCRSRKETEIPMELSAALPLKTDVNDOLDKASV 114

Qy 244 LTVLIAFTKETSABVAQAQKLMVQAADLLSAIHNSLHHGIAQNDTTKGDHPIMMGFEPL 303
 Db 115 MRITIAFLK-----MLQFVSLRCDNDIDKQDIETAEDEQEVKPKLVGTEWD 165

Qy 304 VN-----QRLPPTFPYAKIKREEMVNYFARLIDRIKTVCVENVNLNLHCILDFFCEFS 359
 Db 166 LNGAEARELLKQTMGDFLLVLSHEGDIYVS-----ENVVEYLGITKIDTGLQOIWEYS 219

Qy 360 EQ-----SPCVLSRSI-----LQTFIVDNKKV----- 382
 Db 220 HQDHAIEIKALSUKRELAKVKDEPOQNSGVSTHHRDLFVRLKCTLTSGRSINIKAS 279

Qy 383 -----FGTHLMQDMVKDALRFDVPPVLSKPCYLYNNHQAQKDCIDSFT-HCVRPFCSLI 436
 Db 280 YKVIHITGLVNVNAKGERLLMAIGRPHP-----SNIEIPLGTSTFLTKH-----SLD 328

Qy 437 QIHGHNRRARQDKLGHILEPATLODEMTYFN-----RAKVDAAHLTMLLKQEPORQ 491
 Db 329 MRFTYVDVDMHLLGY-----SPKDLDTSLFSCQHGADSERLMATFKSVLSKQGQETS 382

Qy 492 HLACIG-----TWVL-----YHNL-----TWIOVLLSGFELELYSMHEYYIYWLSEF 536
 Db 383 RYRELKYGKVCWILSQATIVYDKLPQSVVCVNVVISNLE-----NKEIY----- 429

Qy 537 LYAWLMTSLRADQSMAEERIMEBQKGRSKTKKKKKVRPLSRITMSQAYQNMOCAG 596
 Db 430 -----SLAQQAASEQKQHQAAETEKEPEKAADPEIIAQETKETVNTPT 474

Qy 597 MPKTMVAFDMGKVRKPKFELDSEQVREHFRAPFNSVMTPPPHYL-----QFKEMSDLN 652
 Db 475 IHTS-----ELQAKPLQLESEKA-----EKTIETKTATIATPPVTTATSTADQIQLPESN 524

Qy 653 KYSPPQSPPELYVAASKHFOQAKMILENIPNDHEVNLKLVAKFNFMVKLLAG-----GH 709
 Db 525 PY-----KQILQALL-----IKRHNHSPG-----PRTITAQLLSSGSL 560

Qy 710 KKEKVPPEFDFSAHKYFPVVKL 732

Db 561 RPEEKRPKSVTASVLRPSPAPPL 583

RESULT 15

US-09-442-100-12
 ; Sequence 12, Application US/09442100
 ; Patent No. 6359193
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, Tian
 ; APPLICANT: Tao, Wufan
 ; APPLICANT: Wang, Weyi
 ; APPLICANT: Zhang, Sheng
 ; APPLICANT: Yu, Wan
 ; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF LATS
 ; TITLE OF INVENTION: GENES AND METHODS BASED THEREON
 ; NUMBER OF SEQUENCES: 16
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036-2711
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/442,100
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/411,111
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Mistrock, S. Leslie
 ; REGISTRATION NUMBER: 18,872
 ; REFERENCE/DOCKET NUMBER: 6523-003
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 790-9090
 ; TELEFAX: (212) 869-9741/8864
 ; TELEX: 66141 PENNIE
 ; INFORMATION FOR SEQ ID NO: 12:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 526 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 US-09-442-100-12

Query Match 2.7%; Score 105; DB 4; Length 526;
 Best Local Similarity 19.0%; Pred. No. 0.16;
 Matches 106; Conservative 88; Mismatches 205; Indels 160; Gaps 26;

Qy 185 ANSVTDLRVTGMLXDVDDMORRVKSTRSGEER-----DPEVELEHQCCLAVFSTR 236
 Db 34 ASNTTKORVAAAKQYIEKHREQMKNLQERR-ERRILLEKKLADADYSEEDQNLLKE-- 90

Qy 237 VKFTFVLLTVLIAFTKETSABVAQAQKLMVQAADLLSAIHNSLHHGIAQNDTTKGDHP 295
 Db 91 -----LEKKETEMRLQRHKMGADDFELLTMIGKGA-----GEPI 126

Qy 296 IMMGFEPLVNOQLPPTFPYR-YA-KIKREEMVNYFARLIDRIKTVCVENVNLNLHCILD 353
 Db 127 CMIGFSEVITQNCREKTTGQVYAMKLLKSEMLR-RQVEHVKAERNLLAEVDSDCIVK 184

Qy 354 FCECFSEQSPCVLSRSLQTTFLVDNKKVFTHLMQD-MVKDALRSEFVDPVLSPKCYL 411
 Db 185 LYFSQDDYLYLWVEYLPQGDMM-----TLLMRKDILTEDEARFVAETVLAIESIH 237

Qy 412 YNNHQAQKDCIDSFVTHCVRPFCSLIQIHGHNRRARQDKLGH-LEEF-----ATLQD 462

```
Db 238 KHNYIHRD-----IKPDNLLL-----DRYGHKLSDFLCKPLDCSTLEE 277
Qy 463 EFMIFYFN-----RAEKVDAALHTMLLKQEPQROHLACLGTVWLYHNLRIQY----- 511
Db 278 KDFSVGNANGSRSDSPAPKRT-----QQEQLEH-----WQKRRMLAYSTVGTGTP 324
Qy 512 -----LLSGFELELYSMHEYIYIYWLSEFLYAML-----MSTLSR-----A 548
Db 325 DYIAPEVLLKKGYGMEC-----DWWSLGAINYEMLVGYPPFYSDDPMSCTCKIVNWK 376
Qy 549 DGSQMAEERIMEEQKG-----RSSKTKKKKKVPPLSPREITMSQAYQNMCMAGM 597
Db 377 NHLKFPPEAKLSPEAKDIIISRLLCNVTERLGSGADEIKVHSWFKGIDWDRIYQMEAA-- 434
Qy 598 FKTWVAEDMDGKVRKPKFELDSQVRYEHREAPFNSVMTPPVH-----YLQPKEMSD-- 650
Db 435 FIPEVNDELDTQ-NFEKFESESHSQSGSRGPPWRKMLSSKDINFVGYTYKNFKVVDYQ 493
Qy 651 -----LNKYSPPPQSPBL 663
Db 494 VPGKVELKKTNTKPKKPTI 512
```

Search completed: August 10, 2004, 21:52:30
Job time : 23 secs